

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 28, 2005, 15:47:38 ; Search time 28553 Seconds  
(without alignments)

11865.614 Million cell updates/sec

Title: US-09-767-041-9

Perfect score: 6992

Sequence: 1 atcgccaaacgaattggca.....gcaacagcaaaagaagctt 6992

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_btg.\*

3: gb\_in.\*

4: gb\_on.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6992	100.0	6992	1	AF155804 Streptoco
2	1195.8	17.1	15401	1	AF118389 Streptoco
3	955.6	13.7	17468	1	AY057915 Streptoco
4	955.6	13.7	17468	6	AX283715 Sequence
5	892.6	12.8	21562	1	STH289861 Streptoco
6	885.6	12.7	14943	1	SPCP514E Streptoco
7	839.6	12.0	14202	1	AX376403 Streptoco
8	798.8	11.4	18239	1	AF349539 Streptoco
9	798.8	11.4	21365	1	AE014245 Streptoco
10	798.8	11.4	34980	6	C0655070 Sequence
11	798.8	11.4	34980	6	AX954530 Sequence
12	743.4	10.6	17596	1	AF355776 Streptoco
13	729.4	10.4	4513	1	AF332894 Streptoco
14	729.4	10.4	4513	1	AF363060 Streptoco
15	729.4	10.4	4522	1	AF332914 Streptoco
16	729.4	10.4	4888	1	AF332893 Streptoco
17	729.4	10.4	4888	1	AF363058 Streptoco
18	727.8	10.4	4888	1	AF363057 Streptoco
19	727.8	10.4	4888	1	AF363059 Streptoco

20	727.8	10.4	6865	6	BD166249	Beta-1,3-
21	727.8	10.4	9987	1	AB050723	Streptoco
22	726.8	10.4	4768	1	AF332896	Streptoco
23	726.8	10.4	4768	1	AF363055	Streptoco
24	726.8	10.4	4768	1	AF363056	Streptoco
25	726.8	10.4	17276	1	AF163833	Streptoco
26	725.2	10.4	4411	1	AF332897	Streptoco
27	725.2	10.4	4411	1	AF332898	Streptoco
28	725.2	10.4	4411	1	AF332899	Streptoco
29	725.2	10.4	4411	1	AF332900	Streptoco
30	725.2	10.4	95596	6	AX602206	Sequence
31	725.2	10.4	128050	1	SAG766849	Streptoco
32	724	10.4	4521	1	AF332895	Streptoco
33	723.6	10.3	13421	1	AY375362	Streptoco
34	721.4	10.3	10206	1	AB017355	Streptoco
35	721.4	10.3	25021	1	AB028896	Streptoco
36	720.4	10.3	2815	1	AF332901	Streptoco
37	720.4	10.3	2815	1	AF332902	Streptoco
38	714	10.2	16448	1	AF337958	Streptoco
39	652.4	9.3	2217	1	AF367973	Streptoco
40	652.4	9.3	2220	1	AF363035	Streptoco
41	652.4	9.3	2226	1	AF363032	Streptoco
42	652.4	9.3	2226	1	AF363033	Streptoco
43	652.4	9.3	2226	1	AF363034	Streptoco
44	650.8	9.3	2217	1	AF363038	Streptoco
45	650.8	9.3	2217	1	AF363039	Streptoco

## ALIGNMENTS

RESULT 1	AF155804	6992 bp	DNA	linear	BCT 19-DEC-1999
LOCUS	AF155804	6992 bp	DNA	linear	BCT 19-DEC-1999
DEFINITION	Streptococcus suis strain 6555 CpsIE (cpsIE) gene, partial cds; Cps2F (cps2F), CpsIG (cpsIG), CpsIH (cpsIH), CpsII (cpsII), and CpsIJ (cpsIJ) genes, complete cds; and CpsIK (cpsIK) gene, partial cds.				
ACCESSION	AF155804				
VERSION	AF155804.1	GI:6601338			
KEYWORDS	Streptococcus suis				
SOURCE	Streptococcus suis				
ORGANISM	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.				
REFERENCE	1 (bases 1 to 6992)				
AUTHORS	Smith,H.E., Damman,M., van der Velde,J., Wagenaar,F., Wisselink,H.J., Stockhofe-Zurwieden,N. and Smits,M.A.				
TITLE	Identification and characterization of the cps locus of Streptococcus suis serotype 2: the capsule protects against phagocytosis and is an important virulence factor				
JOURNAL	Infect. Immun. 67 (4), 1750-1756 (1999)				
MEDLINE	99184998				
PUBMED	10085014				
REFERENCE	2 (bases 1 to 6992)				
AUTHORS	Smith,H.E., Veenbergen,V., van der Velde,J., Damman,M., Wisselink,H.J. and Smits,M.A.				
TITLE	The cps genes of Streptococcus suis serotypes 1, 2, and 9: development of rapid serotype-specific PCR assays				
JOURNAL	J. Clin. Microbiol. 37 (10), 3146-3152 (1999)				
MEDLINE	99419130				
PUBMED	10488168				
REFERENCE	3 (bases 1 to 6992)				
AUTHORS	Smith,H.E., Veenbergen,V., Van der Velde,J., Damman,M., Wisselink,H.J. and Smits,M.A.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-JUN-1999) Bacteriology, Animal Science and Health, P.O. Box 65, Lelystad 8200 AB, The Netherlands				
FEATURES	Location/Qualifiers				
source	1. .6992				
	/organism="Streptococcus suis"				
	/mol_type="genomic DNA"				
	/strain="6555"				



Qy 421 ATATGCAAGTTTATTTGAATCACAATAA CAAATTCAAAAAATCTTGTGCAATGGTAG 480  
Db 421 ATATGCAAGTTTATTTGAATCACAATAA CAAATTCAAAAAATCTTGTGCAATGGTAG 480  
Qy 481 TTTTAGGTACAGAAATAGATAAAATTAATTTATCATACCGCTCTATTATTCTGTGGAAG 540  
Db 481 TTTTAGGTACAGAAATAGATAAAATTAATTTATCATACCGCTCTATTATTCTGTGGAAG 540  
Qy 541 AAGCTATAGAGTTTTCACAAAGGGAAGTGGTCGACCAAGCTCTTTATAAATCTACCAAGTG 600  
Db 541 AAGCTATAGAGTTTTCACAAAGGGAAGTGGTCGACCAAGCTCTTTATAAATCTACCAAGTG 600  
Qy 601 AGTTTTAGAGCTAAAGCAATTCGTTTCAGATTTTGGAGTTTGGTAGTTATGATGAAGCG 660  
Db 601 AGTTTTAGAGCTAAAGCAATTCGTTTCAGATTTTGGAGTTTGGTAGTTATGATGAAGCG 660  
Qy 661 TTGATATTAATTCANTCGGTTTTACTCGGTTTGAATAA CAAAAAATCCAACTGCTAGGTG 720  
Db 661 TTGATATTAATTCANTCGGTTTTACTCGGTTTGAATAA CAAAAAATCCAACTGCTAGGTG 720  
Qy 721 ACCATAGCATTTGAATCTTTTCCACAAATTTTATAAGCCTAGTCATATCATGATGAAC 780  
Db 721 ACCATAGCATTTGAATCTTTTCCACAAATTTTATAAGCCTAGTCATATCATGATGAAC 780  
Qy 781 GACTTTTGGATATATCTCGAGCGGTAGTCGGGTTAAATTTATTTGGTATAGTTCTATTT 840  
Db 781 GACTTTTGGATATATCTCGAGCGGTAGTCGGGTTAAATTTATTTGGTATAGTTCTATTT 840  
Qy 841 TGTAGTTCCAATTTATTCGTAGAGATGGTGACCGGCTATTTTTCGCAGAAAACGAGTTG 900  
Db 841 TGTAGTTCCAATTTATTCGTAGAGATGGTGACCGGCTATTTTTCGCAGAAAACGAGTTG 900  
Qy 901 GACAGAATGACGCAATTTTACATTTCTACAAAGTTTCGATCGATGATGTCGCTGAGG 960  
Db 901 GACAGAATGACGCAATTTTACATTTCTACAAAGTTTCGATCGATGATGTCGCTGAGG 960  
Qy 961 AGCCCAAAAAAGACTTGTCTAGCCAAAACCCAGATGCAAGGGTGGGTATGTTTTAAAAATGG 1020  
Db 961 AGCCCAAAAAAGACTTGTCTAGCCAAAACCCAGATGCAAGGGTGGGTATGTTTTAAAAATGG 1020  
Qy 1021 GAAAAACGATCCTAGAAATTTACTCCAATTTGGACATTTCTATCGCAAAAAACAGTTTAGACG 1080  
Db 1021 GAAAAACGATCCTAGAAATTTACTCCAATTTGGACATTTCTATCGCAAAAAACAGTTTAGACG 1080  
Qy 1081 AGTTTACCACAGTTTATAATGTTTTAAATTTGCGATATGAGTCTAGTTGCTACAGTCCAC 1140  
Db 1081 AGTTTACCACAGTTTATAATGTTTTAAATTTGCGATATGAGTCTAGTTGCTACAGTCCAC 1140  
Qy 1141 CTACAGTTTGATGAATTTGAAAAATATATCTCCTGGTCAAAAGAGACGATTTGAGTTTAAAC 1200  
Db 1141 CTACAGTTTGATGAATTTGAAAAATATATCTCCTGGTCAAAAGAGACGATTTGAGTTTAAAC 1200  
Qy 1201 CAGGATTTACAGGTCTCTGGCAGGTAGTGGTCTGATGTAATAATCACAGACTTCGACGACG 1260  
Db 1201 CAGGATTTACAGGTCTCTGGCAGGTAGTGGTCTGATGTAATAATCACAGACTTCGACGACG 1260  
Qy 1261 TAGTTCGGTTCGACTTAGCATACATTTGATTAATTTGGACTATCTGGTTCAGATTTAAAAATTT 1320  
Db 1261 TAGTTCGGTTCGACTTAGCATACATTTGATTAATTTGGACTATCTGGTTCAGATTTAAAAATTT 1320  
Qy 1321 TATTAAAGACAGTGAAGTGTGTTGTTGAGAGGGAAGTAAGTAAAGATATATGAAG 1380  
Db 1321 TATTAAAGACAGTGAAGTGTGTTGTTGAGAGGGAAGTAAGTAAAGATATATGAAG 1380  
Qy 1381 TTTGTTTGGTTCGTTCTTCAGGGGACATTTGACTCAGTCTGATTTGTTTAAAAACCGTTTT 1440  
Db 1381 TTTGTTTGGTTCGTTCTTCAGGGGACATTTGACTCAGTCTGATTTGTTTAAAAACCGTTTT 1440  
Qy 1441 GGAAGGAAGAAGAGCTTTTGGGTAAACATTTGATAAAGAGGATGCAAGAGTCTTTTGA 1500  
Db 1441 GGAAGGAAGAAGAGCTTTTGGGTAAACATTTGATAAAGAGGATGCAAGAGTCTTTTGA 1500  
Qy 1501 AGAATGAAAAAATGATCCATGTTTACTTTTCCAAACAAATCGCAATCTCATTAATTTAGTGA 1560

Db 1501 AGAATGAAAAAATGATCCATGTTTACTTTTCCAAACAAATCGCAATCTCATTAATTTAGTGA 1560  
Qy 1561 AAAATACCTTTCTTAGCTTTCAAAAAATTTTACGTGATGAGAAACCGAGATGTTATTTTTCAT 1620  
Db 1561 AAAATACCTTTCTTAGCTTTCAAAAAATTTTACGTGATGAGAAACCGAGATGTTATTTTTCAT 1620  
Qy 1621 CTGTGCGGCGGTTGCTGTCCTCTTTTATCATCGGAAAACTATTTTGGAGCAAGACGA 1680  
Db 1621 CTGTGCGGCGGTTGCTGTCCTCTTTTATCATCGGAAAACTATTTTGGAGCAAGACGA 1680  
Qy 1681 TTTTATTTGAAGTATTTGATCGAGTTAATAAATCTACATTAACCTGGAATACTAGTTTATC 1740  
Db 1681 TTTTATTTGAAGTATTTGATCGAGTTAATAAATCTACATTAACCTGGAATACTAGTTTATC 1740  
Qy 1741 CCGTAAACAGATATTTTATTTTCTCAGTGGGAAGAAATGAAGAAGTATATCTCTAAATCTA 1800  
Db 1741 CCGTAAACAGATATTTTATTTTCTCAGTGGGAAGAAATGAAGAAGTATATCTCTAAATCTA 1800  
Qy 1801 TTAACTTTGGGAGTATTTTAAATGATTTTAACTAGTAGGAACTCATGAACACAGTT 1860  
Db 1801 TTAACTTTGGGAGTATTTTAAATGATTTTAACTAGTAGGAACTCATGAACACAGTT 1860  
Qy 1861 TAATCGATTTGATAAAGAGATTTGATTTTGAATAAATGGAAGTATATAACCGAGAAAT 1920  
Db 1861 TAATCGATTTGATAAAGAGATTTGATTTTGAATAAATGGAAGTATATAACCGAGAAAT 1920  
Qy 1921 ATTTATTTCAAACAGGATATTTCTGACTATATTTCCAGAATATTTGCAAGTATAAAAAATTTCT 1980  
Db 1921 ATTTATTTCAAACAGGATATTTCTGACTATATTTCCAGAATATTTGCAAGTATAAAAAATTTCT 1980  
Qy 1981 CAGTTACAAAGAAATGGAACAATATATAA CAAATCAGAAGTAGTTATTTGCCACGGAGG 2040  
Db 1981 CAGTTACAAAGAAATGGAACAATATATAA CAAATCAGAAGTAGTTATTTGCCACGGAGG 2040  
Qy 2041 CCCCCTACTTTTATGAAATTCATTTCCAAAGGAAAAAACAATTTATTTGTTTCTTAGACA 2100  
Db 2041 CCCCCTACTTTTATGAAATTCATTTCCAAAGGAAAAAACAATTTATTTGTTTCTTAGACA 2100  
Qy 2101 AAAAAAGTATGGTGCAACATGTAATAATGATCATCAAGTAGAGTTTGTAAAGAAATTTTACA 2160  
Db 2101 AAAAAAGTATGGTGCAACATGTAATAATGATCATCAAGTAGAGTTTGTAAAGAAATTTTACA 2160  
Qy 2161 AGATAATAATAATTTTATTTATAGAAAAATATAGATGATTTGTTGAAAAAATTTATGAAAT 2220  
Db 2161 AGATAATAATAATTTTATTTATAGAAAAATATAGATGATTTGTTGAAAAAATTTATGAAAT 2220  
Qy 2221 TTCTAAGCAAACTAACTTTTACATCAATAATAATTTTGTGTAAGATTTAAACAAAT 2280  
Db 2221 TTCTAAGCAAACTAACTTTTACATCAATAATAATTTTGTGTAAGATTTAAACAAAT 2280  
Qy 2281 AGTTGAAAAATTTTAACTGAGGATCAAGAAAAATGAATAATAAAGAAATGATATTTGATAA 2340  
Db 2281 AGTTGAAAAATTTTAACTGAGGATCAAGAAAAATGAATAATAAAGAAATGATATTTGATAA 2340  
Qy 2341 TGGCTTATCATAAATTTTCTCAGATTTTACCTGGAGAGGATACAGATATTTATCATCTTCT 2400  
Db 2341 TGGCTTATCATAAATTTTCTCAGATTTTACCTGGAGAGGATACAGATATTTATCATCTTCT 2400  
Qy 2401 CTCAGGAGATTCACACCATTTAGTTCCCTTCAGAAATACCTGTATAATTTTAAATATTC 2460  
Db 2401 CTCAGGAGATTCACACCATTTAGTTCCCTTCAGAAATACCTGTATAATTTTAAATATTC 2460  
Qy 2461 TCAGATTTTATATGTTGAAATTTTACAAAGATGACCAAAATATAAAGAAATAGATATA 2520  
Db 2461 TCAGATTTTATATGTTGAAATTTTACAAAGATGACCAAAATATAAAGAAATAGATATA 2520  
Qy 2521 TGAACGAGTTTAAATGTTTACAGATTTATTTCTTAAATATATCAGAAAAAATCTATTGATAAT 2580  
Db 2521 TGAACGAGTTTAAATGTTTACAGATTTATTTCTTAAATATATCAGAAAAAATCTATTGATAAT 2580  
Qy 2581 ACTGTTTAGAAATTTTATTAAGAATGTATCGAGCTTTTGAATCTATTTTACAAAGATTTGTT 2640  
Db 2581 ACTGTTTAGAAATTTTATTAAGAATGTATCGAGCTTTTGAATCTATTTTACAAAGATTTGTT

Db 2581 ACTGTTTAGAATTTTATTAAAGATGATCGAGCTTTTGAATPACTATTTTACAAAGATTGTT 2640  
Qy 2641 GTTTATTGATAGATAAATAAACATGGTCTAAGAATAAGATTGGTCTTAATATGGGTTTCG 2700  
Db 2641 GTTTATTGATAGATAAATAAACATGGTCTAAGAATAAGATTGGTCTTAATATGGGTTTCG 2700  
Qy 2701 CTTCCACATGATTTTGTGGCAATTCCTTTTATCAAAATGAAACGAAACAGCTTTATTTATT 2760  
Db 2701 CTTCCACATGATTTTGTGGCAATTCCTTTTATCAAAATGAAACGAAACAGCTTTATTTATT 2760  
Qy 2761 AAGTAAATCTAAATCTCAGATGAATCTATTTATACAGACAATTTATAGAAAATATGAATTT 2820  
Db 2761 AAGTAAATCTAAATCTCAGATGAATCTATTTATACAGACAATTTATAGAAAATATGAATTT 2820  
Qy 2821 TCAAAATAGATTATCTAAATATGGAATTTTAAAGATATATAAAGTGGAAAAATCAACATCT 2880  
Db 2821 TCAAAATAGATTATCTAAATATGGAATTTTAAAGATATATAAAGTGGAAAAATCAACATCT 2880  
Qy 2881 TCTCCTATTCCTTTACAGATGATTCCTATTGATGAATTTGCTAAATGCAAGNAATTTAGGT 2940  
Db 2881 TCTCCTATTCCTTTACAGATGATTCCTATTGATGAATTTGCTAAATGCAAGNAATTTAGGT 2940  
Qy 2941 TTTTATTTCGTAAGAAAGTTAAATAATAGAAAATAAAATCTAAATTTAAAGAAATTTACT 3000  
Db 2941 TTTTATTTCGTAAGAAAGTTAAATAATAGAAAATAAAATCTAAATTTAAAGAAATTTACT 3000  
Qy 3001 AAAAAATAAATAGTTGATTTTGGAGAGTAAGTATGTTTAAATTTTAAATATGACC 3060  
Db 3001 AAAAAATAAATAGTTGATTTTGGAGAGTAAGTATGTTTAAATTTTAAATATGACC 3060  
Qy 3061 CGGAATATTTTATTTTAAAGTACTTCGGTGGATTTATTTTCCAGAGCAAAAGTATG 3120  
Db 3061 CGGAATATTTTATTTTAAAGTACTTCGGTGGATTTATTTTCCAGAGCAAAAGTATG 3120  
Qy 3121 TATTTTATTAATTTTATGAAATTTAAATTTTATTTTATATATAAAATTTTGAATACTAAGC 3180  
Db 3121 TATTTTATTAATTTTATGAAATTTAAATTTTATTTTATATATAAAATTTTGAATACTAAGC 3180  
Qy 3181 TAATATTAATAAATGAAATTTTATGTTTTTATATGCGTCTATATATGTTTTGTTTTGAG 3240  
Db 3181 TAATATTAATAAATGAAATTTTATGTTTTTATATGCGTCTATATATGTTTTGTTTTGAG 3240  
Qy 3241 TAGTCACAAGTATGTTTTGTTGAAATAAATTTGGAAGATTTTTCGAGATTTTACTGCTC 3300  
Db 3241 TAGTCACAAGTATGTTTTGTTGAAATAAATTTGGAAGATTTTTCGAGATTTTACTGCTC 3300  
Qy 3301 CCATAATTTGGATTTATGCAATATGTAATTAATTTTGTATTCATTTTATAAATATTTGATT 3360  
Db 3301 CCATAATTTGGATTTATGCAATATGTAATTAATTTTGTATTCATTTTATAAATATTTGATT 3360  
Qy 3361 ATAAAAAATTAATAAATAGTATCTTTTTTATGTTTTTATGTTTTTATAGGTATATCTGCAT 3420  
Db 3361 ATAAAAAATTAATAAATAGTATCTTTTTTATGTTTTTATGTTTTTATAGGTATATCTGCAT 3420  
Qy 3421 TGTATATTTTCAAAATGGAAGATATGTAATTTTATAGACACACCTTATAGGACTAG 3480  
Db 3421 TGTATATTTTCAAAATGGAAGATATGTAATTTTATAGACACACCTTATAGGACTAG 3480  
Qy 3481 ACTATCTTTATAACAGCGGTCAAAACAGAGTTGGTGGCTTTATGAACTATCCTACGTTAA 3540  
Db 3481 ACTATCTTTATAACAGCGGTCAAAACAGAGTTGGTGGCTTTATGAACTATCCTACGTTAA 3540  
Qy 3541 ATACCACTACAAATTAATAGTTTCAATTCGGTTAAATCTTTTGAACCTTATAAAAAATAAATGC 3600  
Db 3541 ATACCACTACAAATTAATAGTTTCAATTCGGTTAAATCTTTTGAACCTTATAAAAAATAAATGC 3600  
Qy 3601 AACAAATTTTTTCTTGCTGCTTTTATACCGATCTATTTAAAGTGGATCGAGATTG 3660  
Db 3601 AACAAATTTTTTCTTGCTGCTTTTATACCGATCTATTTAAAGTGGATCGAGATTG 3660  
Qy 3661 GTAGTTTATCGTAGCAATATTAATATATGCTTTGTTATCGAGATATATAGGTGGAAAAAT 3720  
Db 3661 GTAGTTTATCGTAGCAATATTAATATATGCTTTGTTATCGAGATATATAGGTGGAAAAAT 3720

Qy 3721 TTGCTTGGATAAAAAAGCTAATAGTAAATATTTGTAATACATACTTATTTTAAATACTG 3780  
Db 3721 TTGCTTGGATAAAAAAGCTAATAGTAAATATTTGTAATACATACTTATTTTAAATACTG 3780  
Qy 3781 AATTGCTTTTACCAATGAAATTTTGGCTGTTTATTAATTTCTAGAGATCAAGTAACCAAGCTA 3840  
Db 3781 AATTGCTTTTACCAATGAAATTTTGGCTGTTTATTAATTTCTAGAGATCAAGTAACCAAGCTA 3840  
Qy 3841 GATTTATTTATTTTCAAGGAAGTATTGATAAAGTATTAGAAAACAATATTTTATTTTGGAT 3900  
Db 3841 GATTTATTTATTTTCAAGGAAGTATTGATAAAGTATTAGAAAACAATATTTTATTTGGAT 3900  
Qy 3901 ATGGAATATCCGAATATTCAGTTACGGGAACTTTGGCTCGGAAGTCATTTCAGGCTATATAT 3960  
Db 3901 ATGGAATATCCGAATATTCAGTTACGGGAACTTTGGCTCGGAAGTCATTTCAGGCTATATAT 3960  
Qy 3961 CATTTTATTAATATCAGGAATAGTTGGGTTGATTTTACTGATGTTTCTTTTTTTTATG 4020  
Db 3961 CATTTTATTAATATCAGGAATAGTTGGGTTGATTTTACTGATGTTTCTTTTTTTTATG 4020  
Qy 4021 TTATAAAAAAGCTTATGGAAGTTAAATGGGGAAAAACAGCACTATTTTATTTTACATCATTAG 4080  
Db 4021 TTATAAAAAAGCTTATGGAAGTTAAATGGGGAAAAACAGCACTATTTTATTTTACATCATTAG 4080  
Qy 4081 CCATATTTTTCATATATGAAAACAATAGATCCGATTTATTAATATATAGTACTATCTCTTT 4140  
Db 4081 CCATATTTTTCATATATGAAAACAATAGATCCGATTTATTAATATATAGTACTATCTCTTT 4140  
Qy 4141 CTTCAATAGGTATTTGGAATAATATAAATTTTAAAAAGGATATGGAGCAAAAAATCAAT 4200  
Db 4141 CTTCAATAGGTATTTGGAATAATATAAATTTTAAAAAGGATATGGAGCAAAAAATCAAT 4200  
Qy 4201 GATTTAATTTTCACTTATTTGACCAATTTTATAATGTCCTCAAGATTTATCTTGATAATGTTATT 4260  
Db 4201 GATTTAATTTTCACTTATTTGACCAATTTTATAATGTCCTCAAGATTTATCTTGATAATGTTATT 4260  
Qy 4261 AACAGTATTTTAAACCAACATATACCTAAATTTAGAGGTTATTTCTCGTAAATGATGGAAGT 4320  
Db 4261 AACAGTATTTTAAACCAACATATACCTAAATTTAGAGGTTATTTCTCGTAAATGATGGAAGT 4320  
Qy 4321 ACTGATGATTTCTGAGAAAAATTTGCTTAACTATATGAAGAACGATGGAAGATTTAAATAT 4380  
Db 4321 ACTGATGATTTCTGAGAAAAATTTGCTTAACTATATGAAGAACGATGGAAGATTTAAATAT 4380  
Qy 4381 TACAAGAAAAATTAATCGCGTCTAGCAGATGCTCGAAATTTTCGACCTAGAACATGCAACA 4440  
Db 4381 TACAAGAAAAATTAATCGCGTCTAGCAGATGCTCGAAATTTTCGACCTAGAACATGCAACA 4440  
Qy 4441 GGTAAATATATTCGCTTTGTCGATTTCTGATGACTATATAGAAAGTTGCAATGTTTCGAGAGA 4500  
Db 4441 GGTAAATATATTCGCTTTGTCGATTTCTGATGACTATATAGAAAGTTGCAATGTTTCGAGAGA 4500  
Qy 4501 ATCSATGATAATAAATCTGAGTATAATCCGATATAGCAGATAGATTTTGTGTTTAGTA 4560  
Db 4501 ATCSATGATAATAAATCTGAGTATAATCCGATATAGCAGATAGATTTTGTGTTTAGTA 4560  
Qy 4561 GACGAAAAACGGGTATACAAAGAAAAAAGAAATAGTAAATTTTCATGTCCTTAAACGAGAGAA 4620  
Db 4561 GACGAAAAACGGGTATACAAAGAAAAAAGAAATAGTAAATTTTCATGTCCTTAAACGAGAGAA 4620  
Qy 4621 GAGACTGTAAAAAGAAATTTTGTGTCAGGATCTAAATATAGAAAAATTAATGTTGCTGCAAGCTT 4680  
Db 4621 GAGACTGTAAAAAGAAATTTTGTGTCAGGATCTAAATATAGAAAAATTAATGTTGCTGCAAGCTT 4680  
Qy 4681 TATTTCAGAGATATTAATAAGATATAAATCCAAATTAATAATAGAAATTAATGTTGCGTCAAGCTT 4740  
Db 4681 TATTTCAGAGATATTAATAAGATATAAATCCAAATTAATAATAGAAATTAATGTTGCGTCAAGCTT 4740  
Qy 4741 GATTTCGCTTTTAAATTTGGAGGTCTTTGAAACAAATGTAACAGCTGTAGTAGTTGACTAGTA 4800  
Db 4741 GATTTCGCTTTTAAATTTGGAGGTCTTTGAAACAAATGTAACAGCTGTAGTAGTTGACTAGTA 4800



Qy	4801	GAATATTATTAATAATATGTCATTCGTAACAGTTCGCTTATTAATACGAAATCTCTATA	4860
Db	4801	GAATATTATTAATAATATGTCATTCGTAACAGTTCGCTTATTAATACGAAATCTCTATA	4860
Qy	4861	AATAATATGTAATAGTACACAGATTCGAGAAATACCCCTTTAAGTTTAAAGAGAGCTTT	4920
Db	4861	AATAATATGTAATAGTACACAGATTCGAGAAATACCCCTTTAAGTTTAAAGAGAGAGCTTT	4920
Qy	4921	AGTCATTATTTTGTATGCAAAAGTTTAAAGAGAGGTTTAAATGCTTTTAAACCAAAATGTAT	4980
Db	4921	AGTCATTATTTTGTATGCAAAAGTTTAAAGAGAGGTTTAAATGCTTTTAAACCAAAATGTAT	4980
Qy	4981	TCAACAGATTTGTTGGATAATAGAGTCTCTGCGCAATATTAAGAGTCTTATCGAAAGAGATA	5040
Db	4981	TCAACAGATTTGTTGGATAATAGAGTCTCTGCGCAATATTAAGAGTCTTATCGAAAGAGATA	5040
Qy	5041	CGTAGATATCCATTTTAAAGCGAAAGATATTTATCAAGAAAGCATTTAGTTACGTTG	5100
Db	5041	CGTAGATATCCATTTTAAAGCGAAAGATATTTATCAAGAAAGCATTTAGTTACGTTG	5100
Qy	5101	TATTTGATGAATTTTCGCCCTAAACTATATGTAATGTTTATTAAGAAATTTCAAAAGCAG	5160
Db	5101	TATTTGATGAATTTTCGCCCTAAACTATATGTAATGTTTATTAAGAAATTTCAAAAGCAG	5160
Qy	5161	TAGAGGTAATAAATGGATAAAATTAGTGTTATGTTCCAGTTTATAATGAGTAAATATTT	5220
Db	5161	TAGAGGTAATAAATGGATAAAATTAGTGTTATGTTCCAGTTTATAATGAGTAAATATTT	5220
Qy	5221	TAAGTAGTTGTATAGAAAGCAATTAATCAAAATTTATAAANAATATAGAAATATTTTGA	5280
Db	5221	TAAGTAGTTGTATAGAAAGCAATTAATCAAAATTTATAAANAATATAGAAATATTTTGA	5280
Qy	5281	TAGATGATGGCTCTGTAGATGATCTGCTCAAAATATGCAAGAAATATCGCAAAAGATA	5340
Db	5281	TAGATGATGGCTCTGTAGATGATCTGCTCAAAATATGCAAGAAATATCGCAAAAGATA	5340
Qy	5341	AAAGAGTAAATAATTTTTTCACTAATCATAGTGGAGTATCAAAATCTAGAAATCATGGAA	5400
Db	5341	AAAGAGTAAATAATTTTTTCACTAATCATAGTGGAGTATCAAAATCTAGAAATCATGGAA	5400
Qy	5401	TAAAGCGGAGTACAGCTGAATATATATGTTGTTGACTCTGATGATGTTGTTGATAGTA	5460
Db	5401	TAAAGCGGAGTACAGCTGAATATATATGTTGTTGACTCTGATGATGTTGTTGATAGTA	5460
Qy	5461	GATTAGTAGAATAATTAATTTTAAATTAATAAAGTAGAAGTATTTATCTGTTGTT	5520
Db	5461	GATTAGTAGAATAATTAATTTTAAATTAATAAAGTAGAAGTATTTATCTGTTGTT	5520
Qy	5521	TGTACGCTACTTTTTCAGAAATATAATAATTTTGAAGTGAATAATCCAAATATTTGATT	5580
Db	5521	TGTACGCTACTTTTTCAGAAATATAATAATTTTGAAGTGAATAATCCAAATATTTGATT	5580
Qy	5581	TTGAAGCAATTAATAACCGTGCAGGACATGGGAGAAAAAATTTTATGAATTTGTATATA	5640
Db	5581	TTGAAGCAATTAATAACCGTGCAGGACATGGGAGAAAAAATTTTATGAATTTGTATATA	5640
Qy	5641	ATAATATTTTTTCTACTCTGTTGTAACATATATAAAGAAAGATACATACAGATCTTT	5700
Db	5641	ATAATATTTTTTCTACTCTGTTGTAACATATATAAAGAAAGATACATACAGATCTTT	5700
Qy	5701	TTCAAGAGAAATCAATGGTTAGGAGAAATTTTACTTTTAAATCTGCATTATTTAAAGATA	5760
Db	5701	TTCAAGAGAAATCAATGGTTAGGAGAAATTTTACTTTTAAATCTGCATTATTTAAAGATA	5760
Qy	5761	TAGATAGAGTTAGTTATTTGACGAAATCTTTTATTTTATAGGAGAGGTATCTAAGTA	5820
Db	5761	TAGATAGAGTTAGTTATTTGACGAAATCTTTTATTTTATAGGAGAGGTATCTAAGTA	5820
Qy	5821	CAGTAAATCTTTTAAAGAGGTTGTTTTCGAAATTCGAAATTTGCAAAAACAGTGA	5880
Db	5821	CAGTAAATCTTTTAAAGAGGTTGTTTTCGAAATTCGAAATTTGCAAAAACAGTGA	5880
Qy	5881	TAGTATTGTTTAAAGCAATATATGTTGAGGATTTTTCAGCTATCAATTTGTTAAAGATACTA	5940

Db	5881	TAGTATTGTTTAAAGCAATATATGTTGAGGATTTTTCAGCTATCAATTTGTTAAAGATACTA	5940
Qy	5941	TACGTTGCGCAAGTATTTTATTAATAGCTTTTAAAGTAAATACGAAAAACAGTCTATTT	6000
Db	5941	TACGTTGCGCAAGTATTTTATTAATAGCTTTTAAAGTAAATACGAAAAACAGTCTATTT	6000
Qy	6001	TTGCAAAATTTTAAATTTTAAAGTCTTTTATATAAAAAATATTTTAACTTTGTTAAAG	6060
Db	6001	TTGCAAAATTTTAAATTTTAAAGTCTTTTATATAAAAAATATTTTAACTTTGTTAAAG	6060
Qy	6061	TATCTAAACAAATTTCTTTGCTTAAAAATTTTTCGTATAAGAAATTTTTCGAAACAAAGTTT	6120
Db	6061	TATCTAAACAAATTTCTTTGCTTAAAAATTTTTCGTATAAGAAATTTTTCGAAACAAAGTTT	6120
Qy	6121	TTAAAAAATATTTATGTTTAAATAGGAAGATATCATGGATACATTTAGTAAAAATTTCTA	6180
Db	6121	TTAAAAAATATTTATGTTTAAATAGGAAGATATCATGGATACATTTAGTAAAAATTTCTA	6180
Qy	6181	TAAATTTGACCTATATAAATGTAGAAAAATATTTATCTAAATGTATAGATAGCAATTTGPA	6240
Db	6181	TAAATTTGACCTATATAAATGTAGAAAAATATTTATCTAAATGTATAGATAGCAATTTGPA	6240
Qy	6241	ATCAGACCTACAAACATATAGAGATTTCTCTGTTGAATGACGGTAGTACGGATTAATTCGG	6300
Db	6241	ATCAGACCTACAAACATATAGAGATTTCTCTGTTGAATGACGGTAGTACGGATTAATTCGG	6300
Qy	6301	AAGAAAAATTTGTTTAGCATATGCGAAGAAAGATAGTCGATTTTATTTTAAAAAGAGA	6360
Db	6301	AAGAAAAATTTGTTTAGCATATGCGAAGAAAGATAGTCGATTTTATTTTAAAAAGAGA	6360
Qy	6361	ACGCGGGCTATCAGATGCCCCGTAATTTATGTCATAAGTTCGCGCAAGGGTGACTTTAG	6420
Db	6361	ACGCGGGCTATCAGATGCCCCGTAATTTATGTCATAAGTTCGCGCAAGGGTGACTTTAG	6420
Qy	6421	CTTTTATAGACTCAGATGATTTTATTTTATTCGCGAGTTTATCTCCAGCTTTTACAGAGCAA	6480
Db	6421	CTTTTATAGACTCAGATGATTTTATTTTATTCGCGAGTTTATCTCCAGCTTTTACAGAGCAA	6480
Qy	6481	TTGAGAGAGAAATGCCCCCTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	6540
Db	6481	TTGAGAGAGAAATGCCCCCTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	6540
Qy	6541	ATTTCTTTAAACAGCAGAGCGCTTCTCTACAAATCAGGCTGTTCTGAGCGGCAAGGAATGTTT	6600
Db	6541	ATTTCTTTAAACAGCAGAGCGCTTCTCTACAAATCAGGCTGTTCTGAGCGGCAAGGAATGTTT	6600
Qy	6601	GTAAGAAAGCTGCTAGAGCGGATGCTCATCGCTTTGTTGCTGCTGCTGCTGCTGCTGCTG	6660
Db	6601	GTAAGAAAGCTGCTAGAGCGGATGCTCATCGCTTTGTTGCTGCTGCTGCTGCTGCTGCTG	6660
Qy	6661	AAAAAGAACTATTTTGAAGATTTTCGATTTGAAAAGGGTGAAGATTCATGAAGATGAATACT	6720
Db	6661	AAAAAGAACTATTTTGAAGATTTTCGATTTGAAAAGGGTGAAGATTCATGAAGATGAATACT	6720
Qy	6721	TCATTTATCGCTCTCTATGAGTTAGAAAAAGTTGCAATAGTTTAAAGAGTCTGTTGACT	6780
Db	6721	TCATTTATCGCTCTCTATGAGTTAGAAAAAGTTGCAATAGTTTAAAGAGTCTGTTGACT	6780
Qy	6781	ATTAATTTGACCGAGAAAAATAGTATCAAACTTCTAGCATGACTGACCATCGCTTCCATT	6840
Db	6781	ATTAATTTGACCGAGAAAAATAGTATCAAACTTCTAGCATGACTGACCATCGCTTCCATT	6840
Qy	6841	GCCTACTGGAATTTCAAAATGAACGAATGACTTCTATGAAGTAGAGGATTAAGAGC	6900
Db	6841	GCCTACTGGAATTTCAAAATGAACGAATGACTTCTATGAAGTAGAGGATTAAGAGC	6900
Qy	6901	TCCTTACTAGAGTGTATGCTTCTTTTTCGCTTTGCTGCTTTGCTTTTTCGCTTTTTCGCT	6960
Db	6901	TCCTTACTAGAGTGTATGCTTCTTTTTCGCTTTGCTGCTTTGCTTTTTCGCTTTTTCGCT	6960
Qy	6961	ATCATTTGCTTTCAGCAAAACAGCAAAAGAGCTTT 6992	

Db	6961	ATCATTGTTGAGCAAAACAGCAAAAGAGCTT	6992
RESULT 2	AF118389	15401 bp	DNA
LOCUS	Streptococcus suis Cps2A (cps2A), Cps2B (cps2B), Cps2C (cps2C), Cps2D (cps2D), Cps2E (cps2E), Cps2F (cps2F), Cps2G (cps2G), Cps2H (cps2H), Cps2I (cps2I), and Cps2J (cps2J) genes, complete cds; Cps2K (cps2K) gene, partial cds; and unknown genes.		
DEFINITION	AF118389		
ACCESSION	AF118389		
VERSION	AF118389.1	GI:4580620	
KEYWORDS	Streptococcus suis		
SOURCE	Streptococcus suis		
ORGANISM	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.		
REFERENCE	1	(bases 1 to 15401)	
AUTHORS	Smith,H.E., Damman,M., van der Velde,J., Wagenaar,F., Wesseling,H.J., Stockhofe-Zurwieden,N. and Smits,M.A.		
TITLE	Identification and characterization of the cps locus of Streptococcus suis serotype 2: the capsule protects against phagocytosis and is an important virulence factor		
JOURNAL	Infect. Immun. 67 (4), 1750-1756 (1999)		
MEDLINE	99184998		
PUBMED	10085014		
REFERENCE	2	(bases 1 to 15401)	
AUTHORS	Smith,H.E., Damman,M., van der Velde,J., Wagenaar,F., Wesseling,H.J., Stockhofe-Zurwieden,N. and Smits,M.A.		
TITLE	Direct Submission		
JOURNAL	Submitted (07-JAN-1999) Department of Bacteriology, DLO-Institute for Animal Science and Health, P.O. Box 65, Lelystad 8200 AB, The Netherlands		
FEATURES	Location/Qualifiers		
source	1. .15401		
	/organism="Streptococcus suis"		
	/mol_type="genomic DNA"		
	/strain="10"		
	/serotype="2"		
	/db_xref="taxon:1307"		
	<1. .721		
	/note="Orf22"		
	/codon_start=2		
	/transl_table=11		
	/product="unknown"		
	/protein_id="AAD24444.1"		
	/db_xref="GI:4580621"		
	/translation="SLDIDHMEVMEASKAAGSACPSQAYQAAPGAEINIIIVTIT GGLSGFNAAARDMVIIEHPNIIHLIDLSAGSGMDLHVQINRLISAGLDFPOV VEATTHYREHKLFLVAKVDNLKNGRLSKLVCTVVGILNIRWGEASAEGLLELQ KARGHKSVTAAPFEMKAGYDGGRIYMAHNNAKFPQPSSELVKASFPFAIDEVAT SGLCSFYABEGGLMGFEVKA"		
	complement (820. .2079)		
	/note="Orf2y"		
	/codon_start=1		
	/transl_table=11		
	/product="unknown"		
	/protein_id="AAD24445.1"		
	/db_xref="GI:4580622"		
	/translation="MKKYQVIIQDILGIBEHFRPKRGKLPISRQLREQVHCSKDTVQ KAMLEKYQNKIYAVEKSGYVILDRDFQDHTCRAQYRLSRITYEDFRICLKESLIG RENVLPNVHQEGLAELISSVQSLMDYHYTKDQOLVITAGSQQALYILTOMETLA KTBILLIENPTYSRMIELIRHQGIPYQTIERNLDGIDLESIFGTQKIKFPYIPR LHPNLSGDIATKTAIVKLAKQYDVYIIIEDDYADPDSHSPLHYLDTDNRVYIK SFTPLPALRIGALISLPNQRIDIFIKHLSIDYDNLINQKALSLYIDNGMFARNQ HLHHIYHAQWNKI KCLEKALNIPYRIPKGSVTFQLSKGILSPSIQHMFGCYFSG QKADFLOIFFEODFADKLEQFVRYNE"		
	2202. .2936		
	/note="Orf2x"		
	/codon_start=1		
	/transl_table=11		
	/product="unknown"		
	/protein_id="AAD24446.1"		
	/db_xref="GI:4580623"		
	/translation="MKIILPNAKEVNTNLENASFYLLSDRSKPVLDALISQDFVKMAA FYKLNKAEALEADRWYRITGOAKTYPAQQLYDGLMYRMDRGIDISKEENYLRDHV RVATYLGIIHPFEPISPHRLDPQGSIKIQNLSKQYRPPYIDQEVGDDSLIISLASS BEFQVSPQIQKRLVKILFMEKAGQLKVHSTISKGRGRLLSLWAKNNIQLSDIQD FKVDGPEYCTSESTANQLTFIRSIM"		
	3041. .4486		
	/gene="cps2A"		
	3041. .4486		
	/gene="cps2A"		
	/function="putative regulatory protein"		
	/codon_start=1		
	/transl_table=11		
	/product="Cps2A"		
	/protein_id="AAD24447.1"		
	/db_xref="GI:4580624"		
	/translation="MKRGSRSSKPKLVNFALLGLYSITLCLFLVTMYRYNILDPR YLNIYITLLVGAVLAGLIMWRKKARIETALLVFLSVITSYGIQMGQVVKFSTRL NNSFSEYEMSLVPANSDITDVROLTSILAPEDQDNITALLDDISKWESTQLAT SGTSTYLTAVQSMNGESQAMVFNQVTFNILENEDQPSKVKKIYSFKYQTQVETAT KQVSGSFNIYISGIDAYGPISTVSRSDVINIYNVRAHKKILLITTPRDSYVAPADG KQYDKLTHAGIYGVNASVHTLUENFYGIDISYINLNFISFLQIDLVLGGIDVMDQ EFTSLGNHYHFPVQVHLNSDQALGFVRYSYSLTGGDNRGKQKVIAALIKKMSTP ENLKNYQAILSGLEGSIQTDLSLETIMSLVNTQLESQTQFTVESQALTGTGSRDLSY AMPGSOLYMEINQDSLEQSKAAIQSVLVEK"		
	4504. .5193		
	/gene="cps2B"		
	4504. .5193		
	/gene="cps2B"		
	/function="putative role in chain length determination"		
	/codon_start=1		
	/transl_table=11		
	/product="Cps2B"		
	/protein_id="AAD24448.1"		
	/db_xref="GI:4580625"		
	/translation="MNNQEVNAIEIDVFLFKTIRWKKFILLITAVLTAGLAFYYSF LVTPOYDSTTRIVVYVSONVEAGAGLTNQELQAGTYLAKDYREIILSDVLTVATELN LKESLAKISVSPVDTRIVSISVRDADPNEARIANSLTRFAVKVVEYTKVSDVTT LGEAVPAEPTPTPTKRNILLGLLAGILATGLVLVNEVLDDRVKRPQDIEEVMGLTL LGIVPDSKKLK"		
	5203. .5880		
	/gene="cps2C"		
	5203. .5880		
	/gene="cps2C"		
	/function="putative role in chain length determination/export"		
	/codon_start=1		
	/transl_table=11		
	/product="Cps2C"		
	/protein_id="AAD24449.1"		
	/db_xref="GI:4580626"		
	/translation="MAMLEIARTKREGVNTKEYFNARINIQISGADIKVGIITSVK SNEGKTTAAALAIAYARGSKTVLVDADIRNSVMPGFFKPIITKITGLTDYLAGTTDL SQGLCDTIDPNLTVIESKVPNPTALQSKNPNLATLRRYDYVIVDCPPLGLVI DAAIIAQKCDAMVAVVEAGNVKCSLLKVKQEQTGTPTPLGLVILNKYDIATBKSEY GNYGKA"		
	5919. .6650		
	/gene="cps2D"		
	5919. .6650		
	/gene="cps2D"		
	/codon_start=1		
	/transl_table=11		
	/product="Cps2D"		
	/protein_id="AAD24450.1"		
	/db_xref="GI:4580627"		
	/translation="MIDIHSHIIFGVDDGPKTIEESLISLSEAYRQGVYIVATSHRR KGMFTEPKIMINFLQKEAVAQVYPERICLGAELYSKDILSKLEKKVPTLNGS CYLLRFSTDPKWEIQEAVNEMTLGLTVPVLAHIERYDALAQSERVEKLIKDGCVT QVNSNVLPALIGERAKFKKTRYPLEQDLVHCVASDMHNLYSRPPFWREAYQLVK KEYGEDRAKALFKKNPLLILKNQVQ"		
	6675. .8054		
	/gene="cps2E"		
	6675. .8054		
	/gene="cps2E"		



```
Db 14781 TCAGATGCCGTAATTATGTCATAGTTCGCCCAAGGCTGACTATTAGCTTTTATAGAC 14840
Qy 6432 TCAGATGATTTTATTCATTTCGAGTTTCATCAACGTTTACAGCAAGCAATTGAGAGAG 6491
Db 14841 TCAGATGATTTTATTCATTTCGAGTTTCATCAACGTTTACAGCAAGCAATTGAGAGAG 14900
Qy 6492 AATGCCCTTGTGGCAGTTGCTGGTTATGATAGGCTAGATGCTTCGGGGCAATTCCTTAACA 6551
Db 14901 AATGCCCTTGTGGCAGTTGCTGGTTATGATAGGCTAGATGCTTCGGGGCAATTCCTTAACA 14960
Qy 6552 GCAGAGCGCTTCCTCAAAATCAGGCTGTTCTGAGCGGCAGGAATGTTGTAAAGAACTG 6611
Db 14961 GCAGAGCGCTTCCTCAAAATCAGGCTGTTCTGAGCGGCAGGAATGTTGTAAAGAACTG 15020
Qy 6612 CTAGAGCGGATGTCATCGCTTTGTGGTGGCCCTGTATTAACCTCTATAAAAAAGAACTA 6671
Db 15021 CTAGAGCGGATGTCATCGCTTTGTGGTGGCCCTGTATTAACCTCTATAAAAAAGAACTA 15080
Qy 6672 TTTCAAGATTTTCCGATTTGAAAAGGTAAGATTCATGAAGATGAATCTTCACTTATCGC 6731
Db 15081 TTTCAAGATTTTCCGATTTGAAAAGGTAAGATTCATGAAGATGAATCTTCACTTATCGC 15140
Qy 6732 TTGCTCTATGAGTTAGAAAAGTTGCAATAGTTAAAGAGTCTTGTACTATTATGTTGAC 6791
Db 15141 TTGCTCTATGAGTTAGAAAAGTTGCAATAGTTAAAGAGTCTTGTACTATTATGTTGAC 15200
Qy 6792 CGAGAAATAGTATCAACAATCTTAGCATGACTGACCATCGCTTCCATGCGCTACTCGAA 6851
Db 15201 CGAGAAATAGTATCAACAATCTTAGCATGACTGACCATCGCTTCCATGCGCTACTCGAA 15260
Qy 6852 TTTCAAATGAACCAATGCGCTCTCTAAGTAGAGGATGAAGAGCTCTTACTAGAG 6911
Db 15261 TTTCAAATGAACCAATGCGCTCTCTAAGTAGAGGATGAAGAGCTCTTACTAGAG 15320
Qy 6912 TGTATCGTTCATTTTAGCTTTGCTGTTGTTGTTGTTTGGCAATATATCATGCTGTTG 6971
Db 15321 TGTATCGTTCATTTTAGCTTTGCTGTTGTTGTTTGGCAATATATCATGCTGTTG 15380
Qy 6972 AGCAAAACAGCAAAAGAGCTT 6992
Db 15381 AGCAAAACAGCAAAAGAGCTT 15401
```

```
RESULT 3
AY057915 17468 bp DNA linear BCT 01-NOV-2002
LOCUS Streptococcus thermophilus eps type III operon, partial sequence.
DEFINITION
ACCESSION AY057915
VERSION AY057915.1 GI:24473733
KEYWORDS
SOURCE Streptococcus thermophilus
ORGANISM Streptococcus thermophilus
REFERENCE 1 (bases 1 to 17468)
AUTHORS Rallu,F., Ehrlich,D.S. and Renault,P.
TITLE Diversity of eps operons in Streptococcus thermophilus
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 17468)
AUTHORS Rallu,F., Ehrlich,D.S. and Renault,P.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-2001) Genetique Microbienne, INRA, Domaine de
Vilvert, Jouy en Josas 78352, France
FEATURES
source
1..17468
/organism="Streptococcus thermophilus"
/mol_type="genomic DNA"
/db_xref="taxon:1308"
<1..17075
/gene="eps type III operon"
<1..492
/gene="deoD"
<1..492
```

```

/gene="deoD"
/codon_start=1
/transl_table=11
/product="DeoD"
/protein_id="AAL23725.1"
/db_xref="GI:24473734"
/translation="RELJVDYGVKKLIIRVCTAGSLNEDVHVRELVLAAQAAATNSNIIR
NWPDYDFQIIASFNLDDKAYHAKNFMGTHVGNVLSDDVPFYSNFEKNIELKGWV
KAVEMFAALYYLAAHQVDALAIMTIISDLNPDDETTAERQNTFTDMKVGLETL
IAD"
828..2288
/gene="eps3A"
828..2288
/gene="eps3A"
/codon_start=1
/transl_table=11
/product="Eps3A"
/protein_id="AAL23726.1"
/db_xref="GI:24473735"
/translation="MSSRINRKKRTGNRSWGMVNVGLTILYAILALVLLFTMFNPNF
LGRFRNIITIGLLVLAISIPLOKTKKSPVTVTVLVIFSLVSLVGIFGFKOMDI
TKRMQTAASEVEMSIIVPKESDIKDVSQLTSVQAPTVDKXNIILMSALKOKKV
DVKVDVASYQRAYDNLKSGSKAMVLSGYSASLESVDNSFNASKITTYIKIKKK
NNSANOVDSKVENIYISGIDTYGSIISTVSRSDVNIIMVNMNTHKLLTTPRDAYVK
IPGKANQVDKLTAGIYGVETSEOTLEDLYGTIDYARINFTSLKLIDQLGGVTV
HNDQATSLHGKDFDPVGDIOQNSEQALGFVREYISLDGDDNRGRNQEKVISAIVNK
LASLKSVNFTSIYNNLQDSVQTNISLDTINALANTQLDSDGSKPTVTTSQAVTGTGSG
QUTSYAMPNPSIYMMKLDNSSVASASQAIKNLMEEK"
2289..3020
/gene="eps3B"
2289..3020
/gene="eps3B"
/codon_start=1
/transl_table=11
/product="Eps3B"
/db_xref="GI:24473736"
/translation="MIDVSHIVPDVDDGPKTLEESLDLIGESYAQGVKIVSTSHR
KGMFETPKIFANFSKABEALYPLDTIYGGELYTLDIVKEKKNLIINMHN
QFALIEFSAKTSWKEIHSGLSNVLRAGVTPVAHIERYDALEENADRVERIINMGCV
QVNSHVLKPLFGDKKVRKVRPFLEKNLVHMVASDMHNLGPRPPFMKDAYEIVK
KNYGSRAKNLFIENPKLTLENQYL"
3029..3721
/gene="eps3C"
3029..3721
/gene="eps3C"
/codon_start=1
/transl_table=11
/product="Eps3C"
/protein_id="AAL23728.1"
/db_xref="GI:24473737"
/translation="MNQDNTKSDSIDVLALLHLKTKLLILFTAFYFAAFSLGTTF
FIQPTVSTTRIVVNAQDNNLSAQDLQAGTYLANDYKEIITSNDVLSEVTKDEKL
NLSEAKLSKWSVNIPTDRLISISVNAKTQDAQTLANKRVESAKIKKVTKVDEV
TWLEAKLPESPSPNKLNLVGLGFLAVVGVLVREILDNRVRRPEDVEDALGM
ALLGIVPTDKI"
3731..4489
/gene="eps3D"
3731..4489
/gene="eps3D"
/codon_start=1
/transl_table=11
/product="eps3D"
/protein_id="AAL23729.1"
/db_xref="GI:24473738"
/translation="MPLLKLVKSNFAKOTERYNAINRNIQFSGAQIKVIATSSVE
AGEKSTSLNLAISFASVGLTLLIDATRNSVFSCTFKSNPEYKGLSNFLSGNADI
NETICQDTSGLDVIASGPPVNPNTSLLQNDNPRHLMNEVARSYDYIIDTPPYGLVLI
DAVIAHQADASALLTAAGKIKRRFVTKAVEQLEQSGSQFLGVNLNKVDMVTVDKIGSY
GSYGSYGSYGYEYRKTQDTEGSHRHRKRG"
4546..5913
/gene="eps3E"
4546..5913
```

```

/gene="eps3E"
/note="undecaprenylphosphate glucosyltransferase"
/codon_start=1
/transl_table=11
/product="Eps3E"
/protein_id="AAL23730.1"
/db_xref="GI:24473739"
/translation="MKKEQETRRIEIGIIQLVWVFAMVASKIPYTEITOGSIVLIG
VHVVSYYISYIENKRYGDELIAIVKCYFIPALIAITFLSPFADGFSISRGLL
YVWISGVYLVTVTVLKYPFRSSYTRKSNKNILLISDQARLDNVLSRMKMDNGRI
SAVCVLDNPFYDPFIKSKPENLIYATHSVVDQVLNLPSEQVKIWDYASPFELMG
IPVNLNALFMSSQGEKHIQOOLGPFKVFVTFSTFYSYGDILAKRFLLDCCALVGLVL
CGIYFLYPLIRKDGGAIPAIQAQYRVGNGRIKFKYKFSRMRVDAEETKKNLMAQNM
SGCMKMDNDRIRIKHFIRKTSDELQPFNVNLKGMSLVGRPTVDEYKYTPE
QKRRLSPFGITGLWQVSGRSEITDFDEVVKLDVAYIDGWTIWRDIIQILLTKIKVVM
KDGAK"
5946..6264
/gene="orf3A"
/pseudo
5946..6264
/note="orf3A"
/pseudo
5946..6264
/note="orf3A; putative rhamnosyltransferase"
/pseudo
/codon_start=1
/transl_table=11
6275..7647
/gene="eps3F"
/pseudo
6275..7647
/note="eps3F; putative undecaprenylphosphate
glycosyltransferase"
/pseudo
/codon_start=1
/transl_table=11
7682..8131
/gene="eps3G"
/pseudo
7682..8131
/note="eps3G"
/codon_start=1
/note="putative galactosyltransferase enhancer"
/transl_table=11
/product="Eps3G"
/protein_id="AAL23731.1"
/db_xref="GI:24473740"
/translation="MKICLVGSSGHLAHLNMLKPFWKDEERFWVTFDKEDARSILKD
EKYSCHPTNRLNKLKNTFLAIKVLKKEKPDVILSSGAVAVPEFYIGKLFGAKT
YVIEVFRIDRKPVTGKLVVYVDKPIVQNEEMKKVYPKAINLGSIF"
8131..8637
/gene="eps3H"
/pseudo
8131..8637
/note="eps3H"
/codon_start=1
/note="putative galactosyltransferase"
/transl_table=11
/product="eps3H"
/protein_id="AAL23732.1"
/db_xref="GI:24473741"
/translation="MIFVTGTHQEPNRLIQEVDHLVETGVIKEEVFIQYCSYIEP
KFCQWSRUISDQSKFMQKADIIITHGGPATFMSAITNGKKPIVWPQKFGHBND
HOVDFARNVAKRMGTIIILVEDVDKLGETIENYSIVAELEYDMGNMNNKPKNLEGIV
VKLLDNVK"
8642..9358
/gene="eps3I"
/pseudo
8642..9358
/note="eps3I"
/codon_start=1
/transl_table=11
/product="Eps3I"
/protein_id="AAL23733.1"
/db_xref="GI:24473742"
/translation="MNKAKTCFLIIAHNNWQKLKLIBCLDSQTHDIPVHVDRKSKDF
DESYFIDSVTKSLKFRFPTVYWGDFSQVWEMFLLEQVAYIEGYDYYHIISGADLPL

```

```

KNNKEIDLFFENKGEKFIILYDNNLTEDNLEISRRTKYHYFLQNYRRRYTENWKNKSPF
TFCBRVSLVLIQIFGVNRKVLNDWQIKYGSQWVSITNELVKIILEHEEKITSIFQYTK
CKNHRDSIFISFNSQVSN"
9374..10474
/gene="eps3J"
9374..10474
/gene="eps3J"
/codon_start=1
/transl_table=11
/product="Eps3J"
/protein_id="AAL23734.1"
/db_xref="GI:24473743"
/translation="WKLIQIPTGGLSPDGINSFIVEVTAMDKSEMDIRVLATNNAEE
STLQKVBSGCEVSIPIYRKQNIYKYPKLPRIYAKEKIDIVHHGSSAISMSVLAA
KLAGCKVRIASHSNTICENQVDQILRRPFDKLYTECFSCGOEGRWLFGRNRFVIVP
NGRSLKYEYNKRLQYRKGLHTPDALVIGHVGRFNEQKNKHYLIQIFYEHKKYP
KSYLVLIQTDVTVKVKQVTEBLENVYIFIGAVNNVPDYLSPADVMLLPSLYEGLP
SVWLEMOISGLPCLISDNIINECKITSLVBPESIEITPTETWAKDVELKIEDRNINKD
KIFKEVRDAGYDIDEDAKKLKELYSLYSPFK"
10496..11166
/gene="eps3K"
/pseudo
10496..11166
/gene="eps3K"
/note="Eps3K; putative glucosyltransferase"
/pseudo

```

	Query Match	13.7%	Score 955.6;	DB 1;	Length 17468;
	Best Local Similarity	65.8%	Pred. No. 6.2e-99;		
	Matches 1484;	Conservative	0;	Mismatches 729;	Indels 41; Gaps 5;
Qy	1	ATCCGCAAAACGAAATGGCATTATTGATATGATAGTAGCAGTAGTGCAATTTCTGCAATCTTAA	60		
Db	6294	ACCGGCAATTTAGACTGGCGCTTTTAGAATTTGATTTGGGGTAGTAGTGCTGTCTCCATTA	6353		
Qy	61	CAAGTCATATACCAATGCTGATTTAAATCGTCTGGAAATTTTATCATATATGATGGTTC	120		
Db	6354	CCACCCACTTGGCCACCCTCAAAGATGAATCAGATAGGGACGCTATTGTAATTTGCCCTGC	6413		
Qy	121	ATTATTTTGCATTTTATATCTCGTATGCGAGTTGAAATTTGAGTATAGAGTAACTCTCA	180		
Db	6414	ATTTTGGGCTTTTATTTCTCAGTATGTCATCAGAAATTTGAGATGAGGCTACTTAA	6473		
Qy	181	TAGAGTTTGAATAAACATTTAACTATAGTATAATTTTGAATTTTCTTACGGCAGTAT	240		
Db	6474	TAGAGTTTGAACGGTCTGCTAGGTATGTTTTCGATTTTTCGATTTACTTCTCATTTT	6533		
Qy	241	CATTTTGTGGAGAAATAATTTCCGACITTTCAAGACGTGGTCCGCTGATTTTCAATTA	300		
Db	6534	CTTTTATTGGGAGATAGTTTATCATTCGCGACGTGGGTGGCATTTATTTACCATTC	6593		
Qy	301	TAAACTTCGTTTTCGTATACCTATTTAACGTAATTTATTAAGCAGTTTAAAGGATAGCTTTC	360		
Db	6594	TTACCTTATCATCGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	6653		
Qy	361	TATTTTCGACATCTATCAAAAAAGACGATTTAAATTAACACGCTCAACGATGGGAAA	420		
Db	6654	TTATGACTGCTGAACAACAGAGAATACTTTAGTCAATTTACCCTACTGAGAGATTCGAT	6713		
Qy	421	ATATGCAAGTTTATTTGAATCAATAAACAAATTCAAAAAATCTTGTTCGATTTGGTAG	480		
Db	6714	TAATGGAGGGGCTCTTTGAATCGGATCAACTTCCTCAACAAATATTTGCGGGTATTGTAG	6773		
Qy	481	TTTTTAGGTACAGAAATAGATAAATAATTAATTTATCATCCGCTCTATTATTTCTGTGAAG	540		
Db	6774	TCATTTGG---AGACGGTGATGTGGCCTTTTCAGAGGGAGTTCCAGTAGTGTCTTTTGATT	6830		
Qy	541	AAGCTATAGAGTTTTCACAAAGGAAGTGGTCGACCACTCTTTTATAAATCTTACCAAGTG	600		
Db	6831	ATGCAATAGAGTTTTCGACCCCATGAGTTGCGACCATGTGTTTATCAACTTACCGAGTG	6890		
Qy	601	AGTTTATAGCGTAAAGCAATTCGTTTCAGATTTTGGTGTGTTAGTATTGATGTAAGCG	660		

Db 6891 AACATTACGATCTCAAAACATCTTGTTCCGATTTTGAAGTCATCGGTATTGAUTGAGTG 6950  
Qy 661 TTGATATTAATTCATTCCGTTTACTCGGTTTGAAACCAAAACCAAAATCAACTGCTAGTG 720  
Db 6951 TAGATATTAACACATCTGATTTTAGGCTTTAAACAAATAAAAAATCAACAGGTTGGAG 7010  
Qy 721 ACCATAGCATTTGAATTTTCCCAAAATTTTATAAGCCCTAGTCATATCATGATGAAC 780  
Db 7011 ACCATAGCATCTGACCTTTTAACCTCAATTAATAACCAATAGCCATATCTTCTAAGC 7070  
Qy 781 GACTTTTGGATPACTCGGACCGTAGTCGGGTTTAATTAATTTGCGTATAGTTTCTATTT 840  
Db 7071 GCATGTTGGATATCTTTGGGCTTTTGATTTGCTTCTTAATTTTGGTCTGCTGGGATTG 7130  
Qy 841 TGTTAGTTCCAAATTAATTCGTAGACGTTGACCGGCTATTTTGGCTCAGAAACGAGTTG 900  
Db 7131 TCTTAGCCCCCATCAITTCGGAAGATGGAGACCGGCCATTTTGTTCAGAAACGAGTAG 7190  
Qy 901 GACGAATGGACGCATATTTACATTTCTACAAGTTTCGATCGATGATGTTGATGCTCAGG 960  
Db 7191 GGAATAACGGACGATCTTTAAATTTTATAATTTCCGTTCTATGTACACTGATCGGAGG 7250  
Qy 961 AGCCCAAAAAAGACTTGTCTAGCCCAAAACCAAGTCAAGGCTGGGTATGTTTAAATGG 1020  
Db 7251 A- ---AAGAAAGAAATGAAGGCACAAAATCAGATGCA- ---CGGTGGAATGTTCAAGAT 7301  
Qy 1021 GAAAAACGATCTAGAAATTTACTCAATTTGGACATTTTCATAGCGAAAAACAAAGTTTAGACG 1080  
Db 7302 GAATAACGATCCAGTATCACTCTATCGGTCTGTTTATTTCG-AAAAACGAGTCTGGACG 7360  
Qy 1081 AGTTACCACAGTTTATAATCTTTTAAATGCGGATATGAGTCTAGTTGGTACAGCTCCAC 1140  
Db 7361 AACTGCTCAGTTTACATATGTTGGTCGCGGATATGAGTCTTGTGGGACTCGCCAC 7420  
Qy 1141 CTACAGTTGATGAATTTGAAAAATATATCTCTCGTCAAAAGAGACGATGAGTTTAAAC 1200  
Db 7421 CAACGTGATGAATTTGAAAAATATACCAAGTCAAAACGTCGCTCAGCTTCAAGC 7480  
Qy 1201 CAGGATATACAGGTCCTGGCAGGTTAGTGTCTAGTATATATACAGACTTCGACGACG 1260  
Db 7481 CAGGAATTTACAGGTCATATGGCAGGCAAGTGGTCGTAGTATATATACCAACTTTGATCAGG 7540  
Qy 1261 TAGTTCGTTGCACTTAGCATACATATGATATTCGACTATCTGTCGATAGATTAATAAATTT 1320  
Db 7541 TGGTAAGACTCGATTTGAATATATCGATACACTGTCGATTTGGTCGATATCAAGATAT 7600  
Qy 1321 TATTAAAGACAGTGAAGTTGTATTTGTTGAGAGAGGGAAGTAAGT----- 1365  
Db 7601 TGTTGAAGACTATTGTCGTGCTTTTAAAGAGAGGGAAGTGTAGTAGAGCCCTCTTCA 7660  
Qy 1366 -----AAAAGTATATGAAGTTTGTGTCGTGCTTCTTTCAGGGGACATTTG 1412  
Db 7661 TTCTTGTATTGGGAGAAAAATGAAATTTGTTTAGTAGGTCATCTGGAGGGCAATTA 7720  
Qy 1413 ACTCACTTGATTTTGTAAACCGTTTGGAGGAAGAAACGCTTTTGGTGAACATTT 1472  
Db 7721 GCACATTTAAATATGTTAAACCAATCTGGAGGATGAAGAACGTTTTTGGTHAACCTTT 7780  
Qy 1473 GATAAAGAGGATCAAGAAGTCTTTTGAAGAAATGAAAAATGTATCCATGTTACTTTCCA 1532  
Db 7781 GATAAAGAGGATCAAGAAGCATCTTAAAGATGAAAAATGTATCTTGTCACTTTCTCT 7840  
Qy 1533 ACAATCGCAATCTCAATTAATTTAGTGAATAATCTTTCTTAGCTTTCAAAATTTTAGT 1592  
Db 7841 ACAATCGAACTTTAAAAATTTAATTTAAAAATATCTTTCTAGCTATTAAGTATTAATA 7900  
Qy 1593 GATGAGAAACAGATGTTATTTTCACTCGTCGCGCGTGTGTCCTCTCTTTTAC 1652  
Db 7901 AAAGAAACCTGAATATTTTCACTCGAGCTCGCGGAGCGGTCCATTTCTTTTAC 7960  
Qy 1653 ATCGAAACATTTTGGAGCAAGACGATTTTATATTTGAAGTATTTGATCGAGTTTAAATA 1712  
Db 7961 ATTGGAACATTTTGGAGCTAAGACGTTTATATTTGAAGTCTTTTGTAGATGAATTTGATAA 8020

Qy 1713 TCTACATTAACTCGAAAACTAGTTTATCCGTTAAACAGATATTTTATTTGTTTCAGTGGAA 1772  
Db 8021 CCAACAGTTTACAGAAAAATTTAGTGTACCTCTGTACCGGATAAGTTTCATTTGTTCAATGGAA 8080  
Qy 1773 GAAATGAAGAAGGTATATCTTAATCTATTAACTTGGGAGTATTTTTTAATGATTTTTG 1832  
Db 8081 GAGATGAAAAAGTTTATCCAAAGCTATTAAATTAGGAGTATCTTTTAATGATTTTTG 8140  
Qy 1833 TAAACAGTAGAACTCATGAAACAAACAGTTTAATCGATTGATAAAGAGATTGATTTATGA 1892  
Db 8141 TAAACAGTAGAAACACACGAACAACTTTTAATCGCTTAATTCAGGAAGTGGATCATTTAG 8200  
Qy 1893 AAAAAATCGAAGTATAACCGACGAATATTATTCAACAGGATATTCTGACTATATTC 1952  
Db 8201 TTGAACAGGAGTGATAAAGAGAGAGGTCCTTAATCAACAGGATATTCTATCTACGAAC 8260  
Qy 1953 CAGAATATTGCAAGTATAAAAAATTTCTCAGTTTACAAAGAAATGGAAACAATATATTAA 2012  
Db 8261 CTAAATTTTGTAGTGTCTCGTTTGTATAGTTTGTGATCAATGAGTAAATTTATGCAA 8320  
Qy 2013 AATCAGAAGTAGTTATTTGCCAGGAGGCCCTACTTTTATGAAATTCATTTTCAAAAG 2072  
Db 8321 AAGCTGATATTAATATCACTCACCGAGGTCAGGACTTTTATGAGTGCCATTACAATG 8380  
Qy 2073 GAAAAACAAATTTATGTTTCTTAGCAAAAAAAGTATGCTGAACATGTAATGATCATC 2132  
Db 8381 GCAAAAGCCAAATTTGTTGTCCTAGACAAGAAAGTTTGGTGAACATGTCAATGACCATC 8440  
Qy 2133 AAGTAGAGTTTGAAGAAGAAATTTTACAAGATAATAATTTTATTTATAGAAAAATATAG 2192  
Db 8441 AAGTAGATTTTGCAGAAATGTAGCAAAAGATGGAACTATCATTTCTGTTGAAGATG 8500  
Qy 2193 ATGATTTGTTGAAAAATTTTGAAGTTTCTAA 2226  
Db 8501 TAGATAAATTTGGGGAGACTATTGAAAAATTACAA 8534

RESULT 4  
AX283715  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
ORIGIN

AX283715 17468 bp DNA linear PAT 20-NOV-2001  
Sequence 9 from Patent WO0179500.  
AX283715  
AX283715.1 GI:17044426  
Streptococcus thermophilus  
Streptococcus thermophilus  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.  
1  
Rallu, F., Besancon-Yoshpe, I., Fremaux, C., Mengaud, J. and Renault, P.  
Streptococcus thermophilus operons involved in exopolysaccharide (e  
ps) synthesis  
Patent: WO 0179500-A 9 25-OCT-2001;  
INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (INRA) (FR) ;  
COMPAGNIE GERVAIS-DANONE (FR) ; Rhodia Chimie (FR)  
Location/Qualifiers  
1. 17468  
/organism="Streptococcus thermophilus"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:1308"

Query Match 13.7%; Score 955.6; DB 6; Length 17468;  
Best Local Similarity 65.8%; Pred. No. 6.2e-99;  
Matches 1484; Conservative 0; Mismatches 729; Indels 41; Gaps 5;  
Qy 1 ATCGCAACGAAATTTGGCATTTTGTATGATGATAGCAGTTGCAATTTCTGCAATCTTAA 60  
Db 6294 ACCGGCAATTTAGCTGGCGCTTTTGAATTTGATTTGGGTAGTAGTGCTGTCTCCATTA 6353  
Qy 61 CAAGTCATATACCAAAATGCTGATTTAAATCGTTCTGGAATTTTATCAATATGATGTTTC 120

Db 6354 CCACCCACTTCCCACTCAAGATGAATCAGATAGGACGCTATTGATTATTGCCCCTGC 6413  
Qy 121 ATTATTTTGCATTTTTTATATCTCGTATGCGAGTTGAAATTTGAGTATAGAGTAACTCTGA 180  
Db 6414 ATTTTGTGGCTTTTATTTCTCACGTATGTCTATCAGATATTTGAGAGTAGAGGCTACTTAA 6473  
Qy 181 TAGAGTTTGAAGAAACATTTAACTATAGTATTAATTTTGCAATTTTCTTACGGCAGTAT 240  
Db 6474 TAGAGTTTGAAGCGGCTCGCTAGGTATGTTTGAATTTTCGATTTTCTCTCATTTTCA 6533  
Qy 241 CATTTTCTGCGAGAAATTAATTTGCGACCTTCAAGACGTGGTCCGTGTAATTTTCACTTAA 300  
Db 6534 CTTTTTATGCGAGATAGTTTATCATATGCGACGTGGTGGCATATTTTACCATTC 6593  
Qy 301 TAAACTCTCGTTTGGTATACCTATTATTAACGTAAATTTAAGCAGTTTAAAGATAGCTTTC 360  
Db 6594 TTACTTTATCATCGTTTATATATCACTGCACCTGTTTCGTAGATACAAATACCTCTTCT 6653  
Qy 361 TATTTTGCACAACTATCAAAAAGACGATTTCTAAATTTACAAACGCTGAAACGATGGGAAA 420  
Db 6654 TTATGACTGCTGAAACACAGAGAATACTTTAGTCAATTTACCACACTACTGAGAGATTGCCAT 6713  
Qy 421 ATATGCAAGTTTATTTGAATCACATAACAAATTTCAAAAATCTTGTTCGATTGGTAG 480  
Db 6714 TAAATGGAGGGCTCTTTGAATCGGATCAACTTCTTACCAAAATATTTGGCGGGTATTTAG 6773  
Qy 481 TTTTAGTACAGAAATAGATAAATTAATTTATCAATACCGCTCTATTATTTCTGTGGAAG 540  
Db 6774 TCATTTGG---AGACGGTATGTGGCTTTCAGAGGGAGTTCCAGTAGTTCCTTTTGATT 6830  
Qy 541 AAGCTATAGAGTTTCAACAGAGGAAGTGTGCGACCGCTTTTATATAAATCTCAAGATG 600  
Db 6831 ATGCAATAGAGTTTGGCAGCCCATGAAGTGTGCGACCATGTGTTTATCAACTTACCGAGTG 6890  
Qy 601 AGTTTATAGAGTAAAGCAATTCGTTTTCAGATTTTGTAGTTTGTAGTATTCATGTAAGC 660  
Db 6891 AACATTAGCTCTCAACATCTGTTTCCGATTTTGAAGTCATGGGTATTCATGTTAGTG 6950  
Qy 661 TTGATATTAATTCATTCGGTTTACTGCGTTTGAAGAAACAAAATAACAACTGCTAGGTG 720  
Db 6951 TAGATATTTACATCTGATTTTATGGGCTTTTAAACAAATTAATAAATCAACAGGTTGGAG 7010  
Qy 721 ACCATAGCATTTGTAATTTTCCAAATTTTATAAGCCTAGTCATATCATGATGAAC 780  
Db 7011 ACCATAGCATCTGACTTTTAACTCCAAATTTACTACAAACATAGCCATCTTTCTAAGC 7070  
Qy 781 GACTTTTGGATATCTCGAGCGGTAGTCGGGTAAATTTATTTGGTATAGTTTCTATT 840  
Db 7071 GCATGTTGGATATCTTTGGGGCTTTGATTTGTTCTTAAATTTTGGTCTGTCGGGATTG 7130  
Qy 841 TGTTAGTTCCAAATTTCTAGAGATGCTGGACCGCTATTTTCTCAGAAACGAGTTG 900  
Db 7131 TCTTAGCCCCCATCATTCGGAAGATGAGAGCCGCGCCATTTTGTTCAGAAACGAGTAG 7190  
Qy 901 GACAGATGAGCGCATATTATCAATTTCTACAAAGTTTCGATCGATGATGTTGATGCTGAGG 960  
Db 7191 GGAAAAACGAGCATCTTTAAATTTTATAAATTTCCGTTCTATGTACACTGATGCGGAG 7250  
Qy 961 AGCGCAAAAAAGACTGTCTAGCCAAAAACAGATGCAAGGGTGGGTATGTTTAAATPG 1020  
Db 7251 A-----AAGAAAGAAATGAAGGCACAAATCAGATGCA-----CGGTGGAATGTTCAAGAT 7301  
Qy 1021 GAAAAACCATCTAGAAATTTCTCAATTTGGACATTTTCATACGCAAAAAACAAGTTTAGACG 1080  
Db 7302 GAATAACCATCCAGTATCACTCTATCTCGTTCGTTTATTCG-AAAAACGAGCTCTGGACG 7360  
Qy 1081 AGTTACCAAGTTTTTATAATGTTTTAAATTTGGCGATATGAGTCTAGTCTGGTACAGCTCCAC 1140  
Db 7361 AACTGCCTCAGTTTTTACAATATGTTTGGTCGGCATATGAGTCTTGTGGACTCGCCAC 7420  
Qy 1141 CTACAGTTGATGAATTTGAAAAATATCTCTCGTCAAAAGAGACGATTTGAGTTTAAAC 1200  
Db 7421 CAACCTGTCGATGAATTTGAAAAATATACCAAGTCAAAACGTCGCGCTCAGCTTCAAGC 7480

Qy 1201 CAGGATTTACAGGTCTCTGGCAGGTTAGTGGTCTAGTAAATATCACAGACTTTCACGACG 1260  
Db 7481 CAGGAAATTACAGGTCTATGGCAGGCAAGTGGTCTAGTAAATATTAACAACTTTCATGAGG 7540  
Qy 1261 TAGTTCGGTTGGACTTAGCATATGATTAATTCGACTATCTGCTCAGATTAATAAATTT 1320  
Db 7541 TGGTAAGACTCGATGTTTGAATATATCGATACTCGTTCGATTTGGTCGGATATCAAGATAT 7600  
Qy 1321 TATTAAAGACAGTCAAGAGTTGTATTTGTGAGAGAGGAAAGTAAGT----- 1365  
Db 7601 TGTTCAGACTATTTGTCTGCTGTTTAAAGAGGAGGAGTAGTAGAGCCCCCTCTTCA 7660  
Qy 1366 -----AAAAGTATATGAAAGTTTGTGTTGTCGTTCTTACAGGGGACATTTG 1412  
Db 7661 TTCTGTATTGGGAGAAAAAATGAAATTTGTTTAGTAGGCTCATCTGAGGGCATTTA 7720  
Qy 1413 ACTCACTTGTATTTGTTTAAACCGTTTGGNAGGAAGAAACGTTTTTGGGTAACTTT 1472  
Db 7721 GCACATTTAAATATGTTTAAACCAATCTCGAAGGATGAAGAACGTTTTTGGGTAACTTT 7780  
Qy 1473 GATAAGAGGATCAAGAAAGTCTTTTGAAGAAATGAAAAATGTATCCATGTTACTTTCCA 1532  
Db 7781 GATAAGAGGATGCAAGAGCATCTTAAAGATGAAAAAATGTATCTTGTCACTTCT 7840  
Qy 1533 ACAATCGCAATCTCATTAATTTAGTAAAAATATCTTTCTTAGCTTTTCAAAATTTACGT 1592  
Db 7841 ACAATCGAACTTTAAAAAATTTAATTTAAAAAATACTTTTCTAGCTATTATAAGTATTA 7900  
Qy 1593 GATGAGAAACAGATGTTATTTTTCATCTGGTGGGCGGTGCTGTCTTCTTCTTTTAC 1652  
Db 7901 AAAGAAAAACCTGATGTCATTTTTCATCTGGAGCTGCGTAGCGGTTTCCATTTCTTTTAC 7960  
Qy 1653 ATCGGAAAACTATTGTCGACAAAGACGATTTATATGAAATTTTGTATGTCGAGTTTAATA 1712  
Db 7961 ATTGTAACCTATTGTCGAGCTTAAGCGTTTATATTTGAAGTCTTTGTAGAAATGATTA 8020  
Qy 1713 TCTACATTAACTGGAATACTAGTTTATCCGTAACAGATATTTTATTGTTTCAAGTGGAA 1772  
Db 8021 CCAACAGTTACAGGAAATTTAGTGTACCTCTGACGGATAGTTTCATTTTCAATGGGAA 8080  
Qy 1773 GAAATGAAGAGGTATATCTTAAATCTATTAACCTTGGGAGTATTTTAAATGATTTTGG 1832  
Db 8081 GAGATGAAAAAGTTTATCCAAAAGCTATTAAATTTAGGAGTATCTTTTAAATGATTTTGG 8140  
Qy 1833 TAAAGTAGGAACCTATGCAACAGTTTAAATCGATTCGATAAAGAGATTCGATTTATGA 1892  
Db 8141 TAAAGTAGGAACACACGAAACACCTTTTAAATCGCTTAAATTCAGGAAGTGATTCATTTAG 8200  
Qy 1893 AAAAAAATGGAAGTATTAACCGACGAAATTTTATTTCAAAACAGGATATTTCTGACTATATTC 1952  
Db 8201 TTGAAACAGGAGTATTAAGGAGAGGCTTTTATTTCAACAGGATATTTCTATCTACGAC 8260  
Qy 1953 CAGAAATATGCAAGTATAAAAAATTTCTCAGTTCAAGAAATGGAACAATATATTAACA 2012  
Db 8261 CTAAATTTTGTGTCAGTGTCTCGTTTGTATAGTTTGTGATCAAAATGAGTAAATTTATGCAA 8320  
Qy 2013 AATCAGAGTATGTTTATTTGCCACGAGGCCGCTACTTTTATGAAATTCATTTTCAAAAG 2072  
Db 8321 AAGCTGATATTTATTCATCTCAGCGAGGTCAGCGACTTTTATGAGTGCCATTTACAAATG 8380  
Qy 2073 GAAAAACAAATTTATTTCTTAGCAAAAAAGTATGTTGAACATCTAAATGATCATC 2132  
Db 8381 GCAAAAACCAATTTGTTGCTAGCAAGAAAAGTTTGGTGAACATGCTCAATGACCAATC 8440  
Qy 2133 AAGTAGAGTTTGTGAAGAGAAATTTTACAGATAATAATTTTATTTATAGAAAAATATAG 2192  
Db 8441 AAGTAGATTTTGCAGAAATCTAGCAAAAAAGATGGAACATCTATTCTGTTGGAAGATG 8500  
Qy 2193 ATGATTTGTTGAAAAAATTTATTCGAAGTTTCTAA 2226  
Db 8501 TAGATAAATTTGGGGGAGACTATTGAAAAATTAACA 8534



RESULT 5	STH289861	21562 bp	DNA	linear	BCT 04-JAN-2002
LOCUS	Streptococcus thermophilus eps locus.				
DEFINITION	Streptococcus thermophilus eps locus.				
ACCESSION	AJ289861				
VERSION	AJ289861.1	GI:18076389			
KEYWORDS	deod gene; Deod protein; Eps 2; Eps 3; Eps 4; eps10 gene; Eps10 protein; eps11 gene; Eps11 protein; eps12 gene; Eps12 protein; eps14 gene; Eps14 protein; eps15 gene; Eps15 protein; eps16 gene; Eps16 protein; eps2 gene; Eps2 protein; eps3 gene; Eps3 protein; eps4 gene; Eps4 protein; eps5 gene; Eps5 protein; eps6 gene; Eps6 protein; eps7 gene; Eps7 protein; eps8 gene; Eps8 protein; eps9 gene; Eps9 protein; epsA gene; EpsA protein; epsB gene; EpsB protein; epsC gene; EpsC protein; epsD gene; EpsD protein; epsE gene; EpsE protein; ORF1; ORF6; pseudogene.				
SOURCE	Streptococcus thermophilus				
ORGANISM	Streptococcus thermophilus				
REFERENCE	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.				
	1				
AUTHORS	Pluvinet,A., Charron-Bourgoin,F., Morel,C. and Decaris,B.				
TITLE	Implication of horizontal transfers in the chimeric structure of the eps locus of Streptococcus thermophilus IP6757				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 21562)				
AUTHORS	Pluvinet,A.				
TITLE	Direct Submission				
JOURNAL	Submitted (18-MAY-2000) Pluvinet A., Laboratoire de Genetique et Microbiologie, Universite Henri Poincare - Faculte des Sciences, B.P. 239, Vandoeuvre-les-Nancy 54506, FRANCE				
FEATURES	Location/Qualifiers				
source	1..21562				
CDS	/organism="Streptococcus thermophilus"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:1308"				
	<1..465				
	/note="ORF1"				
	/codon_start=1				
	/transl_table=11				
	/product="hypothetical protein"				
	/protein_id="CAC82000.1"				
	/db_xref="GI:18076390"				
gene	/db_xref="UniProt/TREMBL:Q8VM12"				
	/translation="DIIVGVCLGAQLLSVAYGKYEHSPEBIGVPLTLTDEGLAD EHIQDFSTLTHGHWGMPLGSDKAVVLATSGYPHQIICFPKHYAFQAHLEFDLE AIDLLIADGEHLYQONKQLDFVQTPEQLNHDYSQMNKKLFAFLDSLTOI"				
	800..1809				
	/gene="deod"				
	800..1510				
	/gene="deod"				
	/codon_start=1				
	/evidence=not_experimental				
	/transl_table=11				
	/product="Deod protein"				
CDS	/protein_id="CAC82001.1"				
	/db_xref="GI:18076391"				
	/db_xref="GOA:Q8VM11"				
	/db_xref="UniProt/TREMBL:Q8VM11"				
	/translation="MSIHSAKQGDIAKILLPGDPLRAKFAIENFLEDAVCFNEVRN MFGYTYGKERISIRNMGTMGMPISIVARELIVDYGKKLIRVGTAGSLNENHVHRE LVLAQAAATNSNIIRNDPOYDFPQIANFLLDKAVHIAKNFGMTTHGVNLSDDVFY SNYPEKNIELGKWGVKAVEMAEAAALYYLAQHQVDALAIMTISDSLVPEDDTTAEER QNTFTDMKVGLETLIAD"				
	1744..1764				
	/gene="deod"				
	1781..1786				
	/gene="deod"				
	1804..1809				
terminator	/gene="deod"				
	1853..3307				
	/gene="epsA"				
-35_signal	/function="putative exopolysaccharide regulation"				
-10_signal					
gene					
CDS					
evidence=not_experimental	/transl_table=11				
	/product="EpsA protein"				
	/protein_id="CAC82002.1"				
	/db_xref="GI:18076392"				
	/db_xref="GOA:Q56038"				
	/db_xref="UniProt/TREMBL:Q56038"				
	/translation="MSSRTNRKQKHTSNMGWNVNGLTILYAILALVLLFTWFNPNF LSRFPLNIIITGLLVLAISIFLOTKKLPVTVTVLVIFFSLVSLVGIFFGKQIMIDI TRNRNLTAAFFSEVEMSIIVPKESDIKDVSQLTSVQAPTQVDKNNIETLMSALKDKKDV KVDVDSYQRAYDNLKSGSKAMVLSGSYASLLESVDNSYASNLKTIYTKIKKKK SNANQVDSRVENIYSGIDTYGPISTVRSQVNIIMVNMNTHKLLITTPRDAYVK IPGGAGDQDKLTHAGIYGVETSEQLEDLYGLKDIYARINFTSEKLIDQLGGTVV HNDQAFQKDFPFGDIQMNSEQALGFVRERNLDGDNDRGNKEKVISAILNKLA SLKSVSNFTSIVNNLQDSQVTNMSLNTINALANTQLESQSGSKFTVTQAVTGTGTGQOL ISYAMPNSSLYMMKLDNLSVESASQAIKKLMEEK"				
	3308..4039				
	/gene="epsB"				
	3308..4039				
evidence=not_experimental	/gene="epsB"				
	/codon_start=1				
	/evidence=not_experimental				
	/transl_table=11				
	/product="EpsB protein"				
	/protein_id="CAC82003.1"				
	/db_xref="GI:18076393"				
	/db_xref="UniProt/TREMBL:Q56039"				
	/translation="MIDVHSHIVFDVDDGPTTLEESLDLIGESYAQGVKIVSTSHRR KGMFETPEDKIPANFKKVAEALYPLDTIYGGELYTSDIVEKLNLIPIRMNT QFALIEFSARTSWEKIHSGLSNVLRAGVTPIVAHIERYDALEENADRPRELINMGCVT QVNSHVLAKPLGFDKDKKRVKRRFFLEKLNLMVHVASDMHNLGPPPPFKMDAYEIVK KNYGSKRAKNLFIENPKLTLENOYL"				
	4048..4740				
evidence=not_experimental	/gene="epsC"				
	4048..4740				
	/gene="epsC"				
	/function="chain length determinant and exporter"				
	/codon_start=1				
	/evidence=not_experimental				
	/transl_table=11				
	/product="EpsC protein"				
	/protein_id="CAC82004.1"				
	/db_xref="GI:18076394"				
evidence=not_experimental	/db_xref="GOA:Q8VM10"				
	/db_xref="UniProt/TREMBL:Q8VM10"				
	/translation="MNQDNTKSEIDVLALLHLTKKLLILFTAFYFAVFSFLGTYY FIQPTYTSTRIYVNVQATDNKNLSAQDLQAGTYLANDYKEIITSNDVLSEVTKEDK NLSEAELSKMVSUNIPTDRLISISVNAKTQDAQTLANKRVASKKIKKVTKVEDV TILAEAKLPESPSPNPKLNVLGAVLGGFLAVVGLVREILDORVRPREDVEDALGM TLGIVPDTDKI"				
	4750..5490				
	/gene="epsD"				
	4750..5490				
	/gene="epsD"				
	/function="chain length determinant and exporter"				
	/codon_start=1				
	/evidence=not_experimental				
/transl_table=11					
/product="Epsd protein"					
evidence=not_experimental	/protein_id="CAC82005.1"				
	/db_xref="GI:18076395"				
	/db_xref="GOA:Q8VM09"				
	/db_xref="UniProt/TREMBL:Q8VM09"				
	/translation="MPLLLKLVKVDPAKTEEYYNAINRTNIOFSQAKMVAIASSVE AGRKSTISVNLAI SPASVGLRTLII DAETRNSVLCTPKSNPEYKLSNLSGNADL NETICQTDISGLDVIASGVPVNPPTSLIQNDNFRHLMVARS CYDYVIIDTPVGLVI DAVIIAQADASLLVTGAKIKRRFTVTKAVEQLVESGSQFLGVVNLNKVDMTVDKYGFI GSYGSYGEYGGKSDQKEGSHRAHRRKG"				
	5547..6914				
	/gene="epsE"				
	pseudo				
	5547..6914				

QY	497	AGATAAAATTAAATTTATCAATTTACCGCTCTATTATTTCTGTGGAAGAAAGCTATAGAGTTTTTC	556
Db	7277	AGTGGCGTTTCCAGAGGGAGTCCCAATAAATTCCTTTT---GATGATGCGATTGACTTTGC	7333
QY	557	AACRAAGGAAGTGGTCCGACCAAGCTTTTATAAAATCTACCAAGTGGAGTGGTTTTTACGCGTAAA	616
Db	7334	GACTCATGAGTTGTGACCAATGTTTATCACTTACCGAGTGAACATTTACGATCTCAA	7393
QY	617	GCAATTCGTTTTCAGATTTTGAGTTGTAGGTATTGATGTAAAGCGTTTGATATTAAATTCATT	676
Db	7394	ACATCTCTGTTCCGATTTTGAAGTCATGGGTATTGATGTGAGTGTAGATATTAACTTATT	7453
QY	677	CGTTTTTACTCGGTTGAAAAACAAAAATCCAACTCCTCTAGGTGACCATAGCATTTGTAAC	736
Db	7454	TGATTTTCAGGGCTTTTAAAAAATAAAAAATCAAAACAAAGTTTGGAGCCATAGTATCGTGAC	7513
QY	737	TTTTTCCACAAATTTTTTATAAGCCTAGTCATATCATGATCAAAACGACTTTTGGATTA	796
Db	7514	TTTTTAACTCCAATTACTACAAACATGCCAATATCTTTCTAAAGCGCATGTGTGATATCTT	7573
QY	797	CGAGCGGTAGTCGGGTTAAATTTTGTGGTATAGTTTTCTATTTTGTAGTTCCAAATTTAT	856
Db	7574	TGGGCTGTGATTGGTCTTCTAATTTGCTGGTGGGATTGTCTTAGCTCTTATCAT	7633
QY	857	TCGTAGAGATGGTGG-ACCGGCTATTTTGTCTCAGAAAACGAGTTTGGACAGAAATGGACGCA	915
Db	7634	TCGGAAGATGGAGAAACCTGCTATTTTCTGTTTCTGAGTGGGATTGTCTTAGCTCTTATCAT	7693
QY	916	TATTTTACATTTCTACAAGTTTTCGATCGATGTATGTTGATGCTGAGGAGCGCCAAAAAGACT	975
Db	7694	TCTTTTAATTTTTTATAAATTCGGTTCTATGTATATAGATGACAGAAAAACGGAAAAAAGA	7753
QY	976	TGCTCAGCCAAAAACAGATGCAAGGTTGGGTATGTTTTTAAATATGGAAAAACCATCTAG	1035
Db	7754	TTATGGCTCAAAACAGATGCA-----GGGTGGTATGTTTTAAATGGATATGACCCAG	7808
QY	1036	AATTACTCAAATTCGACATTTTATACGCAAAAAACAGTTTTTGTAGACGAGTTTACACAGTTTT	1095
Db	7809	GATTACACCAATTTGGTCAGTTTATCCG-AAAAACAAGTCTAGATGAATTTGCCACAGTTTT	7867
QY	1096	ATAATGTTTTTAATTTGGCGATATGAGTCTAGTTGGTACACGTCACCTACAGTTTGTATGAAT	1155
Db	7868	ATAATGTACTAGTTTGGCGATATGAGTCTGGTAGGTACTCGTCCACCGACAGTGGATGAAT	7927
QY	1156	TTGAAAAATATACCTCTGGTCAAAAGAGACGATTGAGTTTAAACACGAGGATTACAGGTC	1215
Db	7928	TTGAAAAGTATACACCAAGTCAAAACGTCGCTTCAGCCAGGAATTTACAGGTC	7987
QY	1216	TCTGGCAGTTTGTGGTGGTATATACAGATTTTTCGACGACGAGTTTTCGGTTTGGACT	1275
Db	7988	TATGGCAGTTAAGTGGTGGTATATACCAACTTTTGTATGAGTGGTAAAACTCGATG	8047
QY	1276	TAGCATACATTTGATAATTTGGAATATGTCGTCAGATATATAAATTTTATTAAGACAGTGA	1335
Db	8048	TTGAATATATCCATAAATCTGGTCCGATTGGTCCGATATCAAGATATTTGTTGAAGACTATCG	8107
QY	1336	AAGTTGTTATGTTGAGAGAGGAGTAAGTAA-AGTATATCAAGAGTTTGTGTTGCTCGG	1393
Db	8108	TCGCGGTTTTTCAAAAAAGAGGGAAGTAGATTATATTATGAAAGTTTGTTTAGTAGG	8167
QY	1394	TTCTTCAGGGGACATTTTGACTCACTTGTATTGTTTAAAAACCGTTTTTGAAGCAAGA	1453
Db	8168	TTCTTCGTTGGGCAATTTGGCACATTTGATATGCTTAAACCTTTTGGAGTGAACAGGA	8227
QY	1454	ACGTTTTTGGGTAAACATTTGATAAGAGGATGCAAGAGTCTTTTGAAGAAATGAAAAAT	1513
Db	8228	CGGCTTTTGGGTACCTTTTATGAAGAAGATGCAAGAAATTTTAAAGATGAGCAGTT	8287
QY	1514	GTATCCATGTTACTTTCACAAATCCCAATCTCATTAATTTTGTAGTGAAGAAATACTTCTT	1573
Db	8288	TTATCCGTCGTTATTTTCTACTACAGAAATTTTAAAGAAATTTTGTAGTGAAGAAATACTT	8347





Qy 436 TTGAATCACATAAACAATTCACAAAAATCTTGTTCATTCGTTAGTTAGTTACAGAAA 495  
Db 5455 TCGTAGAATCTGATGAACCTTTTCATGGAACCTGGTAGCAGTAAGTGTGTTGATAAATCTG 5514  
Qy 496 TAGATAAAATTAATTTATCAATACCGCTCTATTTCTGTGGAAGAAAGCTATAGAGTTTT 555  
Db 5515 ATTTTCAACATGATAAAATACCTGTAATGAAA- - - - -GGAATAAATTAATGAATTTG 5568  
Qy 556 CAACAGGGAAGTGGTCGACACAGCTCTTTATAAATCTACCAAGTGAGTTTTTACAGCTAA 615  
Db 5569 CAACGCATGAAGTTGTGATGAGGTGTGTGTCATCTCTCAGGAGAGAGCTATGATATG 5628  
Qy 616 AGCAATTCGTTTCAGATTTTCAGTTGTGTAGTATGATGTAAGCTGTGATATTAATTCAT 675  
Db 5629 GAGAAATTAICTTAAGTTTGAACAATGGGATAGATGTAACTGTAAATCTTAAATGCAT 5688  
Qy 676 TCGGTTTTTACTGCGTTGAAAACAAAAATCCAACTGCTAGGTGACCATAGCATTTGTA 735  
Db 5689 TTAATAAGAAATTTGGGTGCTAATAACAAATTCATGAGATAGTAGGATTGAATCTAGTCA 5748  
Qy 736 CTTTTTCACAAATTTTAAGCCTAGTCATATCATGATGAACGACTTTTGGATATAC 795  
Db 5749 CTTTCTCTACAAATTTTATAAACTAGTCATGTGATTTCAAAGAGAAATCTCGATATTT 5808  
Qy 796 TCGGAGCGGTAGTCGGTTAATTTTGTGTATAGTTTCTATTTCTAGTTCCCAATTA 855  
Db 5809 GTGTGCGCAATTTGGGCTTATCTTTTGTCTATAGCTAGTCTAGTTTGTTCCTATGA 5868  
Qy 856 TTCTAGAGATGTTGGACCGGCTATTTTGTCTCAGAAACGAGTTGGACAGAAATGGACGCA 915  
Db 5869 TTCTGTAAGATGGCGGACAGCTATTTTGTCTCAACTCGTATAGGGAATAATGGTCGAC 5928  
Qy 916 TATTACATCTCAAGTTTCGATCGATGTATGTTGATGCTGAGGAGCGGAAAAAGACT 975  
Db 5929 ATTTTACCTTTTATAAATTCGGTTCGATGCGGATCGATGCTGAAGTTTATCAAGAACAGT 5988  
Qy 976 TGCTCAGCAGAACACAGATGCAAGGTGGGTATGTTTAAATGCGGAACAGATCCTAG 1035  
Db 5989 TGATGGATCAAAATACAGTGGGAGGTGGTATGTTT- - - - -AAGATGACAAATGATCTCG 6043  
Qy 1036 AATTACTCCAAATTTGGACATTTTCATACGCAAAAAACAAGTTTAGACAGTTACCAAGTTTT 1095  
Db 6044 TGTTACAAAAATTCGTCGCTTTATTCG- - - - -TAAACAGTTTAGATGAGTACCCAGTTTT 6102  
Qy 1096 ATATGTTTTTAATTTGGCGATATGATGCTAGTTGTGATACAGTCCACCTACAGTTGATGAAT 1155  
Db 6103 GGAATGCTTTATAGGAGATATGAGTTTGTGGGAAACACGTCACCTACAGTAGACGAGT 6162  
Qy 1156 TTGAAAAATATACCTCTGTTCAAAAGAGAGATGAGTTTAAACAGGATACAGGTC 1215  
Db 6163 ATGATCAGTATCTCCAGAACAGAACGTCGACTCAGCTTTAAACCTGGTATTAACAGTT 6222  
Qy 1216 TCTGGCAGGTTAGTGGTCTAGTAAATATCAGACACTTCGACGAGTAGTTCCGTTGGACT 1275  
Db 6223 TATGGCAGGTTAGCGCGCTAGTAAATAACCGATTTTGACGATGTTGTAATAATAGATG 6282  
Qy 1276 TAGCATATGATAAATTTGATATCTGGTCAGATATTAATAATTTATTAAGACAGTGA 1335  
Db 6283 TGGCTTATTTGATAAATTTGGAACAATCTCGAAGATATTTGAAATTTTTCCTTAAAACTGTA 6342  
Qy 1336 AAGTTGTATTTGTCAGAGAGGAGTAAGTAAAGATATATGAAGTTTGTGTCGGTT 1395  
Db 6343 AAGTTGTATTTATGAGGAATGGAGCGGAGTAGGG- - - - -ATGAAAGTATGCTAGTTGGAT 6399  
Qy 1396 CTTCAGGCGGACATTTGACTCAGTTGTATTTGTTTAAACCGTTTTTGGAAAGGAAGAAC 1455  
Db 6400 CTAGTGGAGGTCATTTAACTCATTGTTATTTTAAACCTTTCTCGAAGATAGAGAA 6459  
Qy 1456 GTTTTGGGTAAATTTGATAAGAGGATGCAAGAGTCTTTTGAAGAAATGAATAATGT 1515  
Db 6460 GATTTGGGTGACTTTTGTATAAAGAGATGCAAGAGTATTTTAGAAAAATGAATATTTT 6519  
Qy 1516 ATCCATGTTACTTTTCCAACAATCGCAATCTCAATTAATTTAGTGAATAATACTTTCTTAG 1575

Db 6520 ATCCATGTTATTAATCTTAATCGTAATCTAAATAATCTTATTAATAAATACGATCTTG 6579  
Qy 1576 CTTTCAAAATTTTACGTGATGAGAAACACAGATGTTATTTTTCATCTGGTGGCGCGTTG 1635  
Db 6580 CTTTCAAAATTTTGAAGAAAGAACGCCCTGATATTTATTCGTCTCATCAGGGCAGCTGTAG 6639  
Qy 1636 CTGTCCTCTCTTTTACATCGGAAACTATTTTGGAGCAAGACGATTTATTTTGAAGTAT 1695  
Db 6640 CAGTTCCTCTCTTTTATCTAGGAAATATTTTGGTCTGAAGACAGTCTATATAGAAGTAT 6699  
Qy 1696 TTGATCGAGTTAATAAATCTACATTAACTCGAAAACTAGTTTATCCCGTGAACAGATATTT 1755  
Db 6700 TTGATAGAAATGATGCTCCGACATGACTGGGAACTTGTTTATCCAGTAAGTCAAGAGT 6759  
Qy 1756 TTATGTTTCAGTCGGAGAAATGAAGAGGTATATCTTAAATCTTATTAACCTGGGGAGTA 1815  
Db 6760 TTATGTTACAAATGGGAAGAAATGAAAAAGTATATCCGAAGCAATCAATCTCGAGGGA 6819  
Qy 1816 TTTTAAATGATTTTGTAAACAGTAGGACTCATGAACAACAGTTTAAATCGATTGATAAA 1875  
Db 6820 TTTTCTAAATGATTTTGTGACAGTAGGTTCTCATGAACAACAGTTTAAATCGACTTATAA 6879  
Qy 1876 AGAGATTTGATTTTGAAGAAAAATGGAAGTATAACCGACGAAATATTTTATCAAAACAGG 1935  
Db 6880 AGAGTTTGATGATTAAGAGTGAGGATTTATTCAGGATGATGTTTATTCANACAGG 6939  
Qy 1936 ATATCTGACTATVATTCAGAAATTTGCAAGTATAAAAAATTTCTCAGTTTCAAAAGAAAT 1995  
Db 6940 ATACTCTAATTAATGTTCCAAAAATTTGTAATCGGAAAAATTAATATCTTATGAAAAAT 6999  
Qy 1996 GGAACAAATATATAAACAATCAGAAAGTATTTTGGCAGGAGGCCCCGCTACTTTTAT 2055  
Db 7000 GAATCAATGATTAAGGAATCAGATATTTATCATTTACCCATGGCGGTCAGCTACGTTTAT 7059  
Qy 2056 GAATTCATATCCAAAGAAAAACAATTTATTTGTTTCTTAGACAAAAAAGTATGTTGA 2115  
Db 7060 GGAGTTTATGCTAAAGGTAAAAATCOAATATTTGTTCCCGGGTAAAAAATTTGTTGA 7119  
Qy 2116 ACATGTAATGATCATCAAGTAGAGTTTGTGAAGAATTTTACAAGATAATATATTTT 2175  
Db 7120 GCATGTAATGATCACCAGATGCAATTTGTAATAATAACGAAAGAAATATACAATTTAAT 7179  
Qy 2176 ATTTTAGAAAAATATAGATGATTTGTTGAAAAAATTTATTTGAAGTT- - - - -TCTAAGCAAC 2232  
Db 7180 AGTTATAGATGATATTTTCAGACTTACATTTAATTTCTTCAATTTTAAAGGACAAACATTT 7239  
Qy 2233 TAACTTTACATCAATAATAATTTTTCGTGAAAGATTTAA- - - - -ACAAATAGTTGAAAA 2289  
Db 7240 TGAACCTTATTCGAATAACGAGAGATTTAATGACGTTTCAATGTGGAATCAGTAACCT 7299  
Qy 2290 ATTTAATGAGGATCAAGAAAAATGAATAAAAAAGATGCATATTTGATAATGGCTTATC 2349  
Db 7300 TTTTAAAGGAAAAATAAATAAATGAAATAGAAATAGAACTCAGTATTTTCTTTTATAAAT 7359  
Qy 2350 ATAAATTTCTCAGATTTTACT 2371  
Db 7360 ACCTTTGGTTTATTTATCTTACT 7381

RESULT 7  
AY376403

LOCUS  
DEFINITION Streptococcus agalactiae strain 7271 Cps7C, Cps7D, Cps7E, Cps7F, Cps7G, Cps7H, Cps7I, Cps7J, Cps7K, Cps7L, NeuB, NeuC, NeuD, and NeuA genes, complete cds.

ACCESSION AY376403

VERSION AY376403.1

GI:38640633

KEYWORDS Streptococcus agalactiae

SOURCE Streptococcus agalactiae

ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.

```
REFERENCE 1 (bases 1 to 14202)
AUTHORS Cieslewicz,M.J., Glusman,G., Chaffin,D., Kasper,D., Madan,A.,
Rodrigues,S., Fahey,J., Wessels,M.R. and Rubens,C.E.
TITLE Evolution of group B Streptococcus capsular polysaccharides
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 14202)
AUTHORS Cieslewicz,M.J., Madan,A., Wessels,M.R., Kasper,D. and Glusman,G.
TITLE Direct Submission
JOURNAL Submitted (28-AUG-2003) Channing Laboratory, Brigham and Women's
Hospital, 181 Longwood Ave, Boston, MA 02115, USA
FEATURES
    source
        1..14202
            /organism="Streptococcus agalactiae"
            /mol_type="genomic DNA"
            /strain="7271"
            /serotype="VII"
            /db_xref="taxon:1311"
        128..820
            /note="capsular polysaccharide chain length regulator"
            /codon_start=1
            /transl_table=11
            /product="Cps7C"
            /protein_id="AAR25945.1"
            /db_xref="GI:38640634"
            /translation="MNKIANTVEININPLNKLKWKKLTITVAIAFATAGLYSIF
            IVTPQYSTRYIYNPNTNNSTAQDLQAGSLANDYKEIITSDVLEKVISSEKL
            NYPSSQLQKITVILKOTRIVISVEDANPKMSQKLANSVREAAVKIKAVTOVEDI
            TLEKGNLPKAPSSPNKKNVLIGFIVGAGLSTVLVIMGILDRVNTTEEDIEKVLG
            TSLGIVPLNKL"
        831..1529
            /note="capsular polysaccharide chain length regulator"
            /codon_start=1
            /transl_table=11
            /product="Cps7D"
            /protein_id="AAR25946.1"
            /db_xref="GI:38640635"
            /translation="MTREIVDSKLQAKTTEYFNAINRTNIQSGKENKILATISVR
            EGEGKSTASTSLAQLAQKFTLLIDADTNSVSGTFFKATGIKGLTNYLSGNADL
            DICETNPLRMLVPSQKVPNPNTALLQNAVFNKIEAIKNIIDYIITDPTGLV
            DAALIANACGFIILVTAQRIKNYVEKAKQMSQSGSKFLGILNKVNESVATYGDY
            GDYNGYKDRKRK"
        1581..2930
            /note="putative glycosyltransferase"
            /codon_start=1
            /transl_table=11
            /product="Cps7E"
            /protein_id="AAR25947.1"
            /db_xref="GI:38640636"
            /translation="MIQTVVYFASLTLTITPNFKNKDLFLVLLIHYIVFLSDF
            YRDFWSRGYLEEFKRWLVKYSPIYIPISLSLFFIFKNSFTWTRLSFFPFIAMNSILLYL
            LNSFLKYVRKYSYAKFSDTKVLIITNKDSLKMTFKNKYDHNVIANCILDSKDCY
            DLKNSLRIINKDALTBELTCLTVDQAFINIPILFGYQIQDIINDIEANGVIYNVN
            VEALSFNIGBKRIQTQEGYSVITYSMKFYKSHLIAKRFIDIMGAILGLLICGIVAI
            FLVPOIRNKGKIPQNSQNRGRIIRFYKFSRMVDAEQIKDLHLLHNQMTGMFK
            LDDPRTIKIKTRKTSIDELPOFYNVLKGMDSLVGTPPTVDEYEKNTQKRRLS
            FKPGITGLWQISGRNNITDPDEIVKLDVQYINENSISWSDIKIILLTKVLLGTGAK"
        2954..3403
            /note="beta-1,4-galactosyltransferase enhancer"
            /codon_start=1
            /transl_table=11
            /product="Cps7F"
            /protein_id="AAR25948.1"
            /db_xref="GI:38640637"
            /translation="MKICLVSSGGHLAHLNPLKSIWEKEDRFWTFDKEDARSILRE
            BIVYHCFPTNRNVLKNTILAPKVLKRPDPVLISSGAAVAPFPFYIKLFGCKT
            IYIVFVRIDIKPTITGKLVPEVTDKFIQWEMKKVYPKAINLOGIF"
        3403..3894
            /note="beta-1,4-galactosyltransferase"
            /codon_start=1
            /transl_table=11
            /product="Cps7G"
            /protein_id="AAR25949.1"
            /db_xref="GI:38640638"
            /translation="MIFVTVTGTHQQFNRLIKEVDRLKGTGAIDQEVFIQTGYSDPEF
            QNCQSKFLSYDDMNCYREAEIVITHGGPATFNNAVSKGKTTIVVPRQEQFGHVVN
            HGVDFNVKVKLIYNFDIIDIIEGLQNLVYEGAMNRPLETEWNSNFIKEFKKILKELCD
            ENQ"
        3881..4975
            /note="capsular polysaccharide repeat unit polymerase"
            /codon_start=1
            /transl_table=11
            /product="Cps7H"
            /protein_id="AAR25950.1"
            /db_xref="GI:38640639"
            /translation="MKINKSLFYIALFVLPNFKSLGLGENSTYKIIVLVAILCAL
            KFLDLSYFERRKLVITLLFVATILNLFVHKVTFILTLIFSLAKDLISLKAPSIIL
            IGSRLGVLLNQIFVKLDVFKYVNFYRDGHFILRSLDGFCHNFNFHFFALFFLY
            IVLNYKRLKLIIVILFTLNLVLYQYTFSTGYIVILFVILYLTNLSLKRPFMKL
            APIQFFLAFTPLSSTIFPNSFVQKLDKLLTGRLNVAHLQQLVDGLTLFGNSPKETS
            VLPDNSYMLLSMYGVLTLCMLIYVYKELIRIVELQQLLFIMSIVLFTSFYPSI
            VMNISMVFGKIFCDGIEPIKEFTIIDI"
        5013..5738
            /note="beta-1,4-glucosyltransferase"
            /codon_start=1
            /transl_table=11
            /product="Cps7M"
            /protein_id="AAR25951.1"
            /db_xref="GI:38640640"
            /translation="MIPKVIHYCWFGGNPLDNLKKYIKTWREQCQPDYEIIEWNERY
            DVTKNIEMRAYTKNFAYYSDVARLDIIVYGGFYLDTDVELLSLDPPLRHVNCFLA
            RELSYDNTGLIIGAIRGOKFIDKNMSVBEESKFLSPDKTCVEITTSLLKGLKSKN
            LIQVIDGYIIPRQYFNPKNLLTGKLDLFDINDTYSIHHYEGSWKPSIFPDSMKIRVR
            LVIDFIFGYTRYRLRLRFLKLQ"
        5782..6744
            /note="beta-1,6-N-glucosaminyltransferase"
            /codon_start=1
            /transl_table=11
            /product="Cps7I"
            /protein_id="AAR25952.1"
            /db_xref="GI:38640641"
            /translation="MKVSLVIVPYNGAPIETTESVMATTYSNGLNINCCDSDTYT
            SVICQTSKDKRKVKYIKENGOSARNLGKLAEGDWIAFLDSDDLCSFPYIBLV
            RVQSETGADIVITGIEDRPNRMEKLTCDKSTYSTLSVEATLTMYDYNNGISPLGKL
            YSKALLRHPYIGKIHEDDTTYKLVASSTRIAPSEKTYVIKQPNSTTQSYREN
            MLYFDAMQNRDPFILRYSKDELIDALNYKTVIGGMVINVKLLEVDKYDYSUREN
            YRKYKSLFNKRVITKEKIKYLSFISSMNLFLEKRIYNYSEG"
        6748..7713
            /note="beta-1,4-galactosyltransferase"
            /codon_start=1
            /transl_table=11
            /product="Cps7J"
            /protein_id="AAR25953.1"
            /db_xref="GI:38640642"
            /translation="MKKISVLIIPVNVQSLFNECIDSVLAQTSNLEIILVDDGSTDT
            SDICDYXKDRILVPHKNGSLDARKKENSRAATGDIYLLSDDDLYLKEDAIEE
            VAGNKGYSIDVLAQVERIGRIINIVILENTIETISPLQAIRNIYNDYDRAIFT
            VHNKLYRLELSTCMYPCVQKGLHEDEFLTKYLKAKNIIPFRYTYNDYIRRESIMT
            WYNSKELHAIIEALQERIEHLKPELVFQSEKALIKTMEVNLTKENDPKYKFPYTL
            RNEVORTILOFIEKQVLLKVKYSLKYLKILNKRKKNMK"
        7710..8666
            /note="alpha-2,3-N-acetylneuraminic acid synthetase"
            /codon_start=1
            /transl_table=11
            /product="Cps7K"
            /protein_id="AAR25954.1"
            /db_xref="GI:38640643"
            /translation="MNYFVCHLYLHLITIKDKENTRIFICDTIPDYCAWIKILN
            NQGIKFLKELDYRQLNQNVNVEVNLVDSYDKHYFERVNTQVILFNDITLMGRYM
            AYLKGNLHLEDGNCQAKLFFGSGVKKRIITYFLKLYVPYGFYKSCUSIEVNSLE
            GLPHDRSKYKELRFLDFTQEQKSLIFKLFKMKPIITAPKSVLLTQPLAQDK
            WYKTATERFOSIQDYDFEGIVHDYRERGNYILKVHPRDAVYSKLPVELLPSNIP
            MEILEMLTGRFEVGITHSTALDPLCTVDRKIILVNLKDIK"
        8663..10063
            /note="capsular polysaccharide repeat unit transporter"
            /codon_start=1
            /transl_table=11
```





Db 3007 TTTGAAATCCATTTTGGGAAAAGAGATAGGTTTTGGGTAAACCTTTTGATAAAGAGATGC 3066  
Qy 1487 AAGAAGTCTTTTGAAGAATGAAAAATGTATCCATGTTTACTTTCCAAACAAATCGCAATCT 1546  
Db 3067 TAGGAGTATCTAAGAGAAGAGATGTTATCAATTCCTTTCTTCCACACACCGTAATGT 3126  
Qy 1547 CATTAATTTAGTGAAGAAATACCTTTCTTAGCTTTTCAAAATTTTACGATGAGAAACCCAGA 1606  
Db 3127 CAAAACCTTGGTAAAAATACTATTTCTAGCTTTTAAAGTCTTTAGAAAGAAAGACCCAGA 3186  
Qy 1607 TGTATTATTTTATCATCTGTGGCGCGTGTGTCCTTTCTTTTACATCGGAAACACTATT 1666  
Db 3187 TGTATCATATCATCTGTGGCGCTGTAGCAGTACCATCTTTTATTTGTTAGTTAAT 3246  
Qy 1667 TGGAGCAAGACGAATTTATTTAGAGTATTTGATCGAGTTAATAAATCTACATTAACCTGG 1726  
Db 3247 TGGTTGTAAGACCAATTTATATAGAGTTTTCGACAGGATAGATAAACCACTTTGACAGG 3306  
Qy 1727 AAAACTAGTTATCCCGTAAACAGATATTTTATTTGTTTCAGTGGGAAAGAAATGAAGAAGT 1786  
Db 3307 AAAATTAGTGTATCTGTAAACAGATAAATTTATTTGTTTCAGTGGGAAAGAAATGAAGAAAGT 3366  
Qy 1787 ATATCCTAAATCTATTAATCTTTGGGAGTATTTTAAATGATTTTGTAAACAGTAGGAAT 1846  
Db 3367 TTATCCTAAGGCAATTAATTTAGGAGGAATTTTAAATGATTTTGTACACAGTGGGACA 3426  
Qy 1847 CATGAACAAAGTTTAAATCGATTGATAAAGAGATTTGATTTATGAAAAAATAATGGAAGT 1906  
Db 3427 CATGAACAGAGTTTCAACCGTCTTATTAAGAGTTGATAGATTAAAGGAGACAGGTCT 3486  
Qy 1907 ATAACCGACGAAATATTTATTTCAACAGGATATTTCTGACTATTTTCCAGAGATATTTGCAAG 1966  
Db 3487 ATTGATCAAGAAGTGTTCATTTCAACCGGTACTCAGACTTTTGAACCTTCAGAAATTTGTCAG 3546  
Qy 1967 TATAAAAAATTTCTCAGTTTCAAGAAATGCAATATATTTAAACAAATCAGAGTAGTT 2026  
Db 3547 TGGTCAAAAATTTCTCTATATGATATGAATGTTTATGATGAGAGAGCTGAAATCGTT 3606  
Qy 2027 ATTTGCCAGGAGGCGCGTACTTTTATGAATTCATTTATCMAAGGAAAAAACAATTA 2086  
Db 3607 ATTACATATGGCGTTCAGCGACGTTTATGAATGACGTTTCTAAGGGGAAAAAACCATA 3666  
Qy 2087 TTGTTTCTTAGCAAAAAAAGTATGGTGAACATATGTAATGATCATCAAGTAGAGTTGTA 2146  
Db 3667 GTAGTTCCAGACAAGACAGTTTCGAGAGCATGTGAATATCATCAGTAGATTTTCTGT 3726  
Qy 2147 AGAAGATTTTACAAGATAATTAATTTATTTATAGAAAAATATAGATGATTTGTTGAA 2206  
Db 3727 AATAAGGTCAAGTTGATCTATAATTTTGATATTTATATAGACATTTGAAGGGTTACAAAAT 3786  
Qy 2207 AAAATTATTTGAAGTTTCTAAGCAAACTAACTTTT---ACATCAAAATAAATTTTGT 2263  
Db 3787 CTAGTTTAAAGGAGCGATGATCGTCGTTTTTTAGAAACCAACATAGTAATTTTAT 3846  
Qy 2264 GAAAGATTAAACCAATAGTTGAAAAATTTTAAATGAGGATCAAGAAATCAATTAATAAAAA 2323  
Db 3847 AAAGAAATTTAAGAGATATTAAGAGGTTGTGATGAATCAATAAAAACTCTTTAT 3906  
Qy 2324 AGATGCAATTTGATAAATGGCTTATATCAATAATTTTCTCAGATTTTACGAGAGGATAC 2383  
Db 3907 TTATATGCAATATTTTATGTTAAATTTTAAATCACTAGGTTTA--GGAGAGGAAAT 3964  
Qy 2384 AGATATTATCATCTCTCAGGAGATGCAACCACTAGTTCTTTCAGATACCTGTAT 2443  
Db 3965 TCAACTTACAAATAGTAATGTTAGTTCGAATCTTTATTTGTGTCGAATAAAAATTTTACTC 4024  
Qy 2444 AATTATTTTAAATTTCTCAGGATTTATATGTTGAATTTTACAAAAGATGAGCAAAAAAT 2503  
Db 4025 GATACCTTTATTTTGAAGAAGAAAACTCGTTACTATCTCTTGTATTTGTCGAGT 4084  
Qy 2504 AAAAGAAATPAGGATATATGAACGAGTTAAATGTTTACAGATTTTCTTAATATATCAGAA 2563  
Db 4085 ATTTTGAATTTTATTTTGTACATAAGGTTTACTTTTATATTAATTTTAAATTTTTCCTA 4144

Qy 2564 AAAACTATGTATAATGTACTGTTTAGAATTTTATTAAGAATGTATCGAGCTTTTGAATAC 2623  
Db 4145 GCAITTTGAAGGATATCTCTCTAAAAAAGCTTTCTCTATAATAATAGGATCGGTATATTG 4204  
Qy 2624 TATTACAAAGATTTGTTGTTTATTGATAGATAAAAAACATGGTCTAAGAAATAGATTG 2683  
Db 4205 GGAGTCTATTAAATCAAAATTTTGTGAAATTTAGATTTTGTAGAAATTAAGATATGTCAT 4264  
Qy 2684 GTTCTAATTTGGTTTTCGCTTCCACATGATTTTGTGGCAATTTCTTTTATCAAAATGAAAAAG 2743  
Db 4265 TTTTATAGAGATGACATTTTCATCTGAGAAGTGAATTTAGGTTTGTGGCATCCCACTTT 4324  
Qy 2744 AAAACAGCTTATTATTTAAAGTAATCTAAATGTCGAGATGAACATAATTATACAGACAAATTA 2803  
Db 4325 ATTCATAATTTTGTGCTTAACCTTTTGTGATATTTGTACTCAATTAATAAACGACTG 4384  
Qy 2804 TAGAAAAATAGAAATTTTCAATAGATTTCTAAATATGGAATTTAAGA-----TATA 2857  
Db 4385 AAGCTTATTGTAATAGTTTATTTTAAACATTAATTTTACTGTATCAATATATCTTTT 4444  
Qy 2858 TAAAGTGGAAAAATCAACATCTTCTCTATTTGCTTTTACAGATGATTTCTATTTGATGAAT 2917  
Db 4445 TCAAGAACAGGATATTTATTTGTAATTTTATTTATTTGCTACTCATTTATCTAACAAAGAT 4504  
Qy 2918 TGCTAAATGCAAGAAATTTAGGTTTTTATTTGCTAGAAAGTTAAAAATGAAAAATAAAT 2977  
Db 4505 AGTCTAATAAAGAGATTTTATGAAATTAGCACCATAATACAGTTTTTTTATTATTAGCA 4564  
Qy 2978 CTAATTTAAAGAAATTTATTTACTAAAAATAAATAGTTTGTGAGAGTAAATGAT 3037  
Db 4565 TTTACTTTTTGAGTCTTCAAAATTTTTCAAATTTTGTTCAAAAATTTAGATAG 4624  
Qy 3038 GTTTAAATTTTAAATATGACCCGGAATTTTATTTTAAAGTACTTCTGCTGATAT 3097  
Db 4625 CTTTTCAGAGTAGTTTAACTGCTCATTTTCACTTTGAGCGTTTGAATCTTTT 4684  
Qy 3098 TTTTATTTCCAGACAAAGATGATTTTATTTTATTTTATTTTAAATTTTAAATTTTATTTCA 3157  
Db 4685 GGAATAGTTTAAAGGAAACAAAGTCTGCTGTTGATATAGTACTCTATGTTA--TTGA 4742  
Qy 3158 TATTAATTTTGAAGAACTAAGCTTAATTTAAAAATAAATGAAATTTTATTTGTTTATTTATG 3217  
Db 4743 GTATGATGTTGTTAGTCTTACCTGTTTGTGATGATAATCTATTTATGTTTACAGTAAG 4802  
Qy 3218 GTCTATATTTATGTTTGTTCAGTAGTCACAGATGTTGTTGAAATAAATTTTGAAG 3277  
Db 4803 AAATAAGAAATAGTTGAACCTTCAATTTACTGCTATTATTAATGCTATAGTATTTATTTAC 4862  
Qy 3278 ATTTATTTGCAGATTTTACTGCTCCCATTAATTTGGATTTTGCATAATGTAATTTATTTAT 3337  
Db 4863 AGAGTTTTTACCAGTATAGTTATGATATTTAGTTGTTAGTTTTCGGAAAAATTTT 4922  
Qy 3338 GTATTCATTTATAAATTTGATTTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3397  
Db 4923 GTGATGTTTGAACCAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 4982  
Qy 3398 AGTTTATTTAGGTATCTGCTATTTATTTTCAAAATCGGAAGATATTTGATTTT 3457  
Db 4983 ATTCGATATAGTAGGATGTCAGGAAACAAATGATACCTAAAGTTTATACATTTATTTGTTG 5042  
Qy 3458 AGACAGACACCTTATAGGACTAGACTATCTTTATTAACAGCGCTCAAAAAAAGGTTGTTGG 3517  
Db 5043 TTTGGAGGAATCCCTTACCTGATTAATTTAAATAAGTACATAAATAAATAAATAAATAAATAA 5102  
Qy 3518 CTTTATGAATCTCTACGTTTAAATACCACTACAAATTTAGTTTCAATTTCCGTTTAACTTT 3577  
Db 5103 TGTCCAGATTACGAAATTTATTTGAATGGAACGAAACGTAATTTATGATTTTACCAAAAAATTT 5162  
Qy 3578 TGCATTTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3637  
Db 5163 TTTATGAGAGAAGCATATATCTAAGAAAAAATTTTGCCTATGTTTCTGATTTATGCAAGGTTG 5222

Qy	3638	CTATTTAAGTGGATCGAGAAATTGGTAGTGTATTATCGCTAGCAATATTAATATATATGCTTTGTT	3697
Db	5223	GATATTTATTACACTTATTGGCGGATTTCTATCTAGACACTGATGTGGAACCTTTTAAAAAGT	5282
Qy	3698	ATGAGATATATAGGTGGAAAAATTTGGTTGGATAAAAAGCTAATAGTATATATTTGTAAT	3757
Db	5283	TTAGCCCTTTGAGAGTTCAATACTGTTTTTTTAGCTAGGAGTTAAAGTTATGATGTGAAT	5342
Qy	3758	AC--TACTTATTTTAAAACTGAAATGCTTTTACCATGAAATTTTGGCTGTTTATAAT	3815
Db	5343	ACAGGACTAATATTGTTGCAATTTAGAGGACAAAATTTATTAAGACAATATGTCGGTA	5402
Qy	3816	TCTAGAGAATCAAGTAACGAAGCTAGATTTATTATTTTATCAAGGAAGTATTTGATAAGTA	3875
Db	5403	TACGAAGAACTCAAATTTTTTAAAGTTTTTGATAAAAAACATGTGTAGAAATCAAGCACGCTTA	5462
Qy	3876	TTAGAAAAACAATATTTTATTGGATATGGAAATATCCGAATATTTCAGTTACGGGAACTTGG	3935
Db	5463	CTTTTAAAAAAGGACTGAAAGTAAAAATATCATACAAGTTATTTGATGGTGCACCTATT	5522
Qy	3936	CTCGGAAGTCAATTCAGGCTATATATCATTTTTTTTATAAAATCAGGAATAGTTGGTTGATT	3995
Db	5523	TATCCAAGACAATACTTTAATCTCAAATAATTTGTGACAGGAAGCTA-----GATTTTC	5577
Qy	3996	TTACTGATGTTCTCTTTTTTTTATGTTTATAAAAAAAGTTATGAGGTTAATGGGAAACA	4055
Db	5578	TAACAATGATACTTATTCTATACATCATTTATGAAGGAAGTTGGAAAAAGCCATCTATT	5637
Qy	4056	GCATATTTTATTATCATCATNTAGCCATATTTTTTCATATATAGMAACAATAGATCCGATT	4115
Db	5638	TTCTCTGATTCGATGAAAAATTA-GAGTAGAAGCTTGTAAATGATTTTATTTTGGATATGGA	5696
Qy	4116	ATTATTATATTAGTACTATTCTTTCTTCAATAGGTATTTGGAATATAATAATTTTAAA	4175
Db	5697	ACATATAGAAGCTGTTGCGTTTTTTTTTAAAGTTAAAAACAGTAGTCGCTATTAACTGAAATA	5756
Qy	4176	AAGGATATGGAGACAAAAAATGAATGATTTAAATTTTCAGTTATTGTACCAATTTATATATGT	4235
Db	5757	TTTTCTTTAAAGAAATGGAGGAACAATGAAGTATCATTAGTCATACCAGTTTATAATGG	5816
Qy	4236	CCAAGATTATCTTGATATAATGTTAAACAGTATTTATTAACCAACAATATCTAATTTAGA	4295
Db	5817	GGCACCTTACATCGAAGAAACATTTAGAAAGTGTATGGCCACACTTATTCAAATCTAGG	5876
Qy	4296	GGTTATTCTCGTAAATGATGGAAGTACTGATGATTTCTGAGAAATTTGCTTTAACTATAT	4355
Db	5877	AAATATTAAATTTGTTGATGCTCAATGATTTACACATCAGTTATATGCCCCAAATAACAC	5936
Qy	4356	GAAGAACGATGGAAAGATTAATAATTTACAGAAAAATTTAATGGCGGCTAGCAGATGCTCG	4415
Db	5937	TTCTTAGAGATAAACCGTGAAATTTATATAAAGAAATGGCGGTCAATCCTCTGCTAG	5996
Qy	4416	AAATTTCCGACTAGAACATGCAACAGGTAAATATATTGCTTTTGTGCGATTTCTGATGACTA	4475
Db	5997	AAATTTAGGAGTCAAACTAGCTGAAGGAGATTGGATTGCTTTCTTTGGAATCCGATGACCT	6056
Qy	4476	TATAGAAGTTGCAATGTTTCGAGAAATGCATGATAATATAACTCAGTATTAATGCCGATAT	4535
Db	6057	CTGCTCACCATTTCTATATAGAAATATTATTAGTAAAGAGTGCAGAGTGAAACTGGTGCTGATTT	6116
Qy	4536	AGCAGAGATAGATTTTTTTGTTTAGTAGACGAAAAACGGGTATACAAAGAAAAAAGAAATAG	4595
Db	6117	GGTGATAACTGGAATCGAAGACCGCCCAACCATAGAATGGAAAAACTTACATGTGATAA	6176
Qy	4596	TAAATTTTCATGCTTTAAACGAGAGAAGACTGTAAAAAGAAATTTTGTGAGGATCTAATAT	4655
Db	6177	GTCAACTTACTCAACACTCAGTGTGGAAGCGACATTAACAGAAATGTATATGATAACTA	6236
Qy	4656	AGAAAAATAATGTTTTGGTCAGCTTTTATTCAGGAGATTTATATAAAGATATAAATTTCCA	4715
Db	6237	CGGTATTTCTCCAATGCGAAAGTTATATCTTAAAGCAATTTATTTAAGACATCATCATACC	6296
Qy	4716	AATTAATAATAGAAAGTATTTGTTGAGGATTTTGTCTTTTTTAAATTTGAGGCTTTTGAAACATGT	4775

[illegible]

Db	7367	ATGTATATCGTATTAGACAGACACTATATGACTGGTTCCTTACAACTTT	7416	
RESULT 8	AF349539	18239 bp	DNA	linear BCT 05-APR-2001
LOCUS	AF349539			
DEFINITION	Streptococcus agalactiae CNCTC 1/82 type V capsular polysaccharide synthesis operon, complete sequence.			
ACCESSION	AF349539			
VERSION	AF349539.1	GI:13549124		
KEYWORDS				
SOURCE	Streptococcus agalactiae			
ORGANISM	Streptococcus agalactiae			
REFERENCE	1	(bases 1 to 18239)		
AUTHORS	McKinnon, K., Chaffin, D.O. and Rubens, C.E.			
TITLE	Streptococcus agalactiae type V polysaccharide synthesis operon complete sequence			
JOURNAL	unpublished			
REFERENCE	2	(bases 1 to 18239)		
AUTHORS	McKinnon, K., Chaffin, D.O. and Rubens, C.E.			
TITLE	Direct Submission			
JOURNAL	Submitted (16-FEB-2001) Department of Infectious Diseases, Immunology and Rheumatology, Children's Hospital and Regional Medical Center, 4800 Sand Point Way NE, Seattle, WA 98105, USA			
FEATURES	Location/Qualifiers			
source	1. 18239			
	/organism="Streptococcus agalactiae"			
	/mol_type="genomic DNA"			
	/strain="CNCTC 1/82"			
	/db_xref="taxon:1311"			
gene	792..18239			
-35_signal	792..797			
-10_signal	818..823			
gene	910..2379			
RBS	910..915			
CDS	922..2379			
	/gene="cpsA"			
	/note="similar to LytR"			
	/codon_start=1			
	/transl_table=11			
	/product="CpsA"			
	/db_xref="GI:13549125"			
	/translation="WNSHRRHOKKHSHTPLRVINFLVIFILLVVLSFLMYRRHFL			
	LAFRLHVIYGVIVLILVLSFLCINKARIFTIILVLASIFVATTLGPKSTIDL			
	TNNLKTSYSEIEMSVVVPKDSKITNIEAVSKLAAPVKNDTSNITDLIEHKSEGI			
	SITPKQDSYDQANRISKDGSQAMLVNAYSLIELSTPDKSQIKTIYTKIKKI			
	NRKNTGKGVFNVIYISGIDFGISTVSRSDVNIIMVTNTHKVLITTPRDYVK			
	IPDGNGYDKLTHAGLVGVTSMKLENLDVYARINFSFKLILDLGGVTV			
	YNDQAFYKHCNFPFVQVTLNSEQALGFVRERYSLGGDNRGKQKVIALLNK			
	LASQSVTKLSIISQLTSVQTNTIINDLNNQLSTGQRTFVESQALTGHGSG			
	ELPSYMPGQQLYMSIDQSSLSNAKSIKNTMBE"			
gene	2372..3116			
RBS	2372..2377			
CDS	2385..3116			
	/gene="cpsB"			
	/note="tyrosine phosphatase"			
	/codon_start=1			
	/transl_table=11			
	/product="CpsB"			
	/db_xref="GI:13549126"			
	/translation="WKICLVGSSCGHLAHLNLLKPIWEKEDRFWTFDKEDARSLIRE			
	EIVYHCFPTNRNVKNLVTILAFVKLRKRPDVISSGAAVAPFFYIIGKLFCKT			
	VYIEVFRMDKPTLTGKLVYPTVDKFIQWEEKVKVPKAINLGGIF"			
	6385..6891			
gene				

gene	3115..3817			
RBS	3115..3120			
CDS	3125..3817			
	/gene="cpsC"			
	/note="putative CPS chain length regulator"			
	/codon_start=1			
	/transl_table=11			
	/product="CpsC"			
	/protein_id="AAK29649.1"			
	/db_xref="GI:13549127"			
	/translation="MNKIANTVEINIFNLKKLWKKKFLITFVAIPATAGLFLYSLF			
	IVPQYTSSTRIYVNPNTNNSITAOLOAGSFLANDYKEIITSTDLKVISSEKL			
	NYPSQLOKITYSILKDTVISVEDANPKSOKLANSVREAAVSKKAVTOVEDI			
	TTLEKGNLKPAPSPFNKKNVLIIGFIVAGLSTVLVIMGLDDRVTEDIEKVLGL			
	TSLGIVPLNKL"			
	3817..4526			
	/gene="cpsD"			
RBS	3817..3822			
CDS	3828..4526			
	/gene="cpsD"			
	/note="CPS chain length regulator/exporter"			
	/codon_start=1			
	/transl_table=11			
	/product="CpsD"			
	/protein_id="AAK29650.1"			
	/db_xref="GI:13549128"			
	/translation="MTRLEIVDSKLROAKTEEYFNARTNIQFSGKENKILAITSVR			
	EGEGKSTASTLSLSLAQAGFKLLIDADTRNSVMSGTFKATGFKGLTYLNSGNADL			
	GDICETNPRLMVVPSGKVPNPPTALLQNAVFNKMLEAKNIIDYIITDPGILGV			
	DAALINACDGFILVTQAGRIKRYNVEKAKEQSGSKFLGILNKVNESVATYGDY			
	GDYNGYKGRDRKRK"			
gene	4564..5927			
RBS	4564..4569			
CDS	4578..5927			
	/gene="cpsE"			
	/note="glucose-1-phosphate transferase"			
	/codon_start=1			
	/transl_table=11			
	/product="CpsE"			
	/protein_id="AAK29651.1"			
	/db_xref="GI:13549129"			
	/translation="MIQTVVVYFSAULTLITPNFKNKOLLFLVLIHYIVFVLSDF			
	YRDFWRYGLEEFKWLKYSFYIFISSLPFIKNSFTTRLSGFTFIAMNSILLYL			
	LNSELYRYKYVYAKFSRDTKVLIITKDSLSKMTFRKXVDHNVIAVCLDSSEKDY			
	DLKNSLRINKDALTSBELTCLTVQAFINIPILFGKYQDIINDIEMAGVIVNV			
	VEALSFNIGEKRIQTEGYSVITYSMKFYKISHLIAKFLDITGAILGLIGLIVAI			
	FLVPIQKDGKPAIFSQNRVGRNGRIFRFYKPSRMVDAEQIKDLLVHNMQTMGLMFK			
	LDDPRITTKGKFKRTSIDELPOFYNVLKGDMSLVGTRPTVDEYKYNSTQKRLS			
	FKPGITGLWISGRNNTDFDEIVKLDVQYINWSIDIKILLTLKVLVLTGAK"			
	5942..6400			
	/gene="cpsF"			
	/note="beta-1,4-galactosyltransferase enhancer"			
	/codon_start=1			
	/transl_table=11			
	/product="CpsF"			
	/protein_id="AAK29652.1"			
	/db_xref="GI:13549130"			
	/translation="MKICLVGSSCGHLAHLNLLKPIWEKEDRFWTFDKEDARSLIRE			
	EIVYHCFPTNRNVKNLVTILAFVKLRKRPDVISSGAAVAPFFYIIGKLFCKT			
	VYIEVFRMDKPTLTGKLVYPTVDKFIQWEEKVKVPKAINLGGIF"			
	6385..6891			
gene				

```
/gene="cpsVG"
6385..6390
/gene="cpsVG"
6400..6891
/gene="cpsVG"
/notes="beta-1,4-galactosyltransferase"
/codon_start=1
/transl_table=11
/product="CpsVG"
/protein_id="AAK29653.1"
/db_xref="GI:13549131"
/translation="MIFVTVGTHQOFNRILKEVDRLKGTCAIDQEVFIQTGYSDFFP
QNCOWSKFLSYDDMNSYMKGAIEIVITHGGPATFNVAYSKGKKTIVVPRQOPGSHVNN
HOYDFVNKVTMYNFDIVDIERLQNVVYEGMMRPFLETNSSNFIEFKVILKELCD
ENQ"
6866..7972
/gene="cpsVH"
6866..6871
/gene="cpsVH"
6878..7972
/gene="cpsVH"
/notes="putative repeating unit polymerase"
/codon_start=1
/transl_table=11
/product="CpsVH"
/protein_id="AAK29654.1"
/db_xref="GI:13549132"
/translation="MKINKSLFYIAIFLVNPFKSLGLGEGNSAYKIIVMLVAILLGI
KFLDLSYFERRKLVIIIFLPIATILNLFVHKVTFILTLIFLALXDI SLKKAFSII
IGRLIGLVNLOI FVKLDLBIKYVNFYRDQOFLRSDGLGFHNPFTNFPALTIPLY
IVLVNKRKPVWMLFILTANLYQVTFSTRGYVILFVILVIVTKNSLIRKVEML
APYQFELLVPTFLSSIFPNSNFVKLDVLLTGRLYHAHLQLVGLTPFGNSFKETS
VLPDMSYMLSLMSLTGVVTFMCFMIIYIYSKIIIIIEQLLLFIMSLIFTESFYPSV
VMNLSWIVFGICDGIPIKKEFTIVNNI"
8001..8736
/gene="cpsVM"
8001..8006
/gene="cpsVM"
8011..8736
/gene="cpsVM"
/notes="putative glycosyltransferase"
/codon_start=1
/transl_table=11
/product="CpsVM"
/protein_id="AAK29655.1"
/db_xref="GI:13549133"
/translation="MIPKVIHYCFWFGNPLDNLKVIKTVREQCPDYIEIWEHNY
DVSXNVFWEAYTKKNFAYSDYARLDIIYTYGGFYLDTDVLLKSLDPLRIHECPFA
REISCDVNTGLIIGAVKHHPKSNMSIYDKSDLTSLNKTCEVETNLLINRGLKKN
IIQKIDITIIPRYFNPKNLLTGKVDCLTSVYSIIHHYEGSWKSSSFISDSLKIRVR
LIIDFLFGYGYRMLLRPLKLKK"
8754..9649
/gene="cpsVN"
8754..8759
/gene="cpsVN"
8762..9649
/gene="cpsVN"
/notes="putative glycosyltransferase"
/codon_start=1
Query Match 11.4%; Score 798.8; DB 1; Length 18239;
Best Local Similarity 48.2%; Pred. No. 2.6e-81;
Matches 3323; Conservative 0; Mismatches 3382; Indels 185; Gaps 31;
QY 16 TGGCAATATTTCATATGATAGTAGTTCGCAATTTCTGCAATCTTAAACAAGTCATATACCAA 75
DB 4573 TAGCGATGATTCACACAGTCTGGTATTATTTCTGCAAGTTTGACATTAACATTAATTA 4632
QY 76 ATGCTGATTTAAATCG---TTCTGGAATTTTATCAATAGATGGTTCATTTATTTGCAAT 132
DB 4633 CTCCCAATTTTAAAGCAATAAGATTTATTTGTTGTTTCTATTCATATATATATTTGCT 4692
QY 133 TTTTATATCTCGTAGCCAGTTCGAAATTTGAGTATAGAGGTAATCTGATAGAGTTTGA 192
```

```
DB 4693 TTTATCTTCTGATTTTACAGAGACTTTTGGAGTCGTGCTATCTTGAAGAGTTTAAA 4752
QY 193 AAACATTTAACTAGATATAATATTTGCAATTTTCTTACGGCAGTATCAATTTTGTGG 252
DB 4753 TGGTATTTGAATACAGCTTTTACTATATTTTTCATATCAAGTTCATTTATTTTATTTTA 4812
QY 253 AGAATAATTTTCGCACATTTCAAGAGCGTGGTCCCGTGATTTTTCACATTAATAAATTCGTTT 312
DB 4813 AAAATCTCTTTTACAACGACACGACTTTCTCTTTTACTTTTATTTGCTATGAATTCGAATTT 4872
QY 313 TGGTATACCTTATTTAACTGAATTTAAGCAGTTTAAAGGATAGCTTCTATTTTTCGACAA 372
DB 4873 TATTATATCTATTGAATTCATTTTAAATAATTTATCGAAATAATTTCTTACGCTAAGTTT 4932
QY 373 TCTATCAAAAAAGAGAGATTTCTAATTTACAACGGCTGAACGATGGGAAAAATATGCAAGTTT 432
DB 4933 CACGAGATACCAAAGTTGTTTGTATAACGAATAAGGATTTCTTTATCAAAAAATGACCTTTA 4992
QY 433 TATTTGNAATCACATAACAATAATTTCAAAAAATCTTTGTTGCAATTTGGTAGTTTTAGGTACAG 492
DB 4993 GGAATAAAT---ACGACCATTAATTTATATCGCTGTCTGTATCTTTGGATTTCTCTGAAAGG 5049
QY 493 AAATAGATAAAATTAATTTTATCATTTACCGCTCTATTATTCTGTGGAGAAAGCTATAGAGT 552
DB 5050 ATTGTTATGATTTGAAACATACTCGTTAAGGATATAAACAAGATGCTCTTACTTCAG 5109
QY 553 TTTTCAACAGGGAAGTGGTCGACCACTCTTTTATAATCTTACCAAGTGGAGTTTGTAGACG 612
DB 5110 AGTTAACTCTGTTAACTGTTGATCAAGCTTTTATTAACATACCACTTGAATTTATTTGGTA 5169
QY 613 TAAAGCAATTC-----GTTTCAGATTTTGAAGTTGTTAGGTATTTGATTAAGCCGTTG 663
DB 5170 AATACCAAAATACAAGATATTATTATTAATGACATTAAGCAATGGGAGTGAATTTGTCAATGTTA 5229
QY 664 ATATTAAATTCATTTCCGTTTACTGCGTTGAAAAACAAAAATTCCAACTGCTAGGTGACC 723
DB 5230 ATGTAGAGCACTTAGCTTTGATAATATAGGAGAAAGCGAATCCAAACTTTTGAAGAT 5289
QY 724 ATAGCAATTTGAACTTTTTCACAAATTTTATAAGCCTAGTCAATATCATGATGAACGAC 783
DB 5290 ATAGTGTATTACATATTTCTATGAAATTTCTATAAATATAGTCCACTTATAGCAAAACGAT 5349
QY 784 TTTTGGATATCTCGAGCGGTAGTCGGGTAAATTTATTTGTTGGTATAGTTTCTATTTTGT 843
DB 5350 TTTTGGATATCACGGGTGCTATTATAGGTTTGTCTATATGTGGCAATTTGTGCAATTTTTC 5409
QY 844 TAGTTCCAAATTTATTCGTAGAGATGGTGACCGGCTATTTTGTCTCAGAAAACGAGTTCGAC 903
DB 5410 TAGTTCCAAATCAGAAAAGATGGTGACCGGCTATCTTTTCTCAAAATAGAGTAGGTC 5469
QY 904 AGAATGGACGCATATTATTCATCTTACAAGTTTTCGATCGATGTATGTTGTGCTGAGAGGC 963
DB 5470 GTAATGGTAGGATTTTATAGATTCTATAAATTCAGATCAATTCGAGTAGATGACGAACAA 5529
QY 964 GCAAAAAAGACTTCTCAGCCAAAACAGATGCAAGGGTGGGTATGTTTAAATGGGAA 1023
DB 5530 TTAAGAAAAGATTTATTAGTTTCACAAATCAAAATGACAGGG-----CTAATGTTTAAAGTTAGA 5584
QY 1024 AAACGATCTTAGAATTTACTCCAATTTGGACATTTTCATACGCAAAAAACAAAGTTTACAGAGT 1083
DB 5585 CGATGATCTTAGAATTTACTAAAATAGGAAAATTTATTCG-AAAAACAAGATAGATAGT 5643
QY 1084 TACCACAGTTTATAATTTTAAATTTGGCGATATGAGTCTAGTTGGTACAGTCCACCTA 1143
DB 5644 TGCTCAATTTCTATAATTTTAAAGGGTATATGAGTTTAGTAGGAACACGCGCTCCCA 5703
QY 1144 CAGTTGATGAATTTGAAAAATATACTCTCGGTCAAAGAGACGATTCAGTTTAAACAG 1203
DB 5704 CAGTTGATGAATTAAGAAAGTAAATTTCAACGCAAGACGACGCTTTAGTTTAAAGCCAG 5763
QY 1204 GGAATTACAGGTCTCTGGCAGGTTAGTGTCTGTAGTATATATACAGACTTCGACACAGCTAG 1263
```



Qy	3397	TAGTTTTATTAGGTATATCTGCATGTTGATATATTATTCAAAATGGGAAAGATATTGTAATTTT	3455
Db	7926	GTATCGAACCTTATAAAAAGGAAATTTACTATTGTGTAATAAT--ATATGACATATTTGCTC	7983
Qy	3457	TAGACACACCTTATAGGACTAGACTATCTTTATAACAGGCGTCAAAACAAGGTTGGTTG	3516
Db	7984	TGATATGGCAGGAGGTAAAGGAATAATGATACCTAAAGTTATACATTAATTTGTTGGTTT	8043
Qy	3517	GCTTTATGAACATCTCCTACGTTTAAATACCACCTACAAATTATAGTTTCAATCCCGTTAACTCT	3576
Db	8044	GGAGGAATCCCTTACCAGATNAATTTAAAGAAATATATAAAACTTCGAGAGAACATGT	8103
Qy	3577	TTGCACCTTATAAAAAATAAATGCACAAATTTTTTTCTTGTGTCTTTCGCTTTTATACCGA	3636
Db	8104	CCGGATTATGAAATTAATTGAATGAAATGAGCATAAATTTATGATGT-----8147	
Qy	3637	TCATTTTAAAGTGGATCGAGAAATGGTAGTTTATCGCTTAGCAATATTAATTTATATGCTTGT	3696
Db	8148	--TAGTAAAAATGTTTTTATAGAGAGAAGCATATACTAAGAAGAAATTTGCTTATGTTTCT	8205
Qy	3697	TATGGAGATATATAGG--TGGAAAAATTTGCTTGGATAAAAAAGCTAATAGTAAATATTGTA	3755
Db	8206	GACTATCGAAGATTGGAATATATTATTAATCTATGCGGGGTTCTATCTAGATATCTGATGTCG	8265
Qy	3756	ATACTACTTATATTTTAAATPACTGAATTCGCTTTACCATGAAATTTTGGCTGTTTATAAT	3815
Db	8266	GAGCTTTTAAAAAGTTTTAGATCCITTCAGGATTCATGAGTTGTTCTTCAGCAAGGAGATT	8325
Qy	3816	TCATAGAAATCAAGTAACGAAGCTAGATTTTATTTATTCAGGAAGTATTGATAAGTA	3875
Db	8326	AGTTGTGATGTGAATACAGGATTAATTAATTTGGCGCTGTTAAAGGACATCACTTTTAAAA	8385
Qy	3876	TTGAAAAACAATATTTTTATTGGGATATGGAAATATCCGAATATTCAGTTACGGGAACCTGG	3935
Db	8386	TCAATATGCTATATATGACAAAGTGATTTAACTTCTCTTAATAAGACATGTGTAGAG	8445
Qy	3936	CTCGGAAGTCAATTCAGGCTATATATCAATTTTTTTTATAAATCAGGAATAGTTGGGTTGATT	3995
Db	8446	GTTTACAACATAATTTATTGATAAACAAGGGCTTAAAGAT--AAGAATTTATTCAAAAGA	8503
Qy	3996	TTACTGATGTTTTCTTTTTTTATGTTATAAAAAAGTTATGGAGTTAATGGGGAACA	4055
Db	8504	TTGATGATATAACAAATATATCCGAGAAATATTTTAAATCCAAAAGAAATTTATTAACAGTA	8563
Qy	4056	GCACATATTTTATTTTACATCATTTAGCCATATTTTTCATATATGAACAATAGATCCGATT	4115
Db	8564	AGGTTGATGTCTGACTAGTGTGTACCTATTCTATACATCAATACGAAGGAAGTTGGAAAA	8623
Qy	4116	ATTATTATATTAGTACTATTTCTTTTCTCAATAGTATTTTGGAAATATAATAATTTTAAA	4175
Db	8624	GTTCTTCATTTATTTCCAGATTCTCTAAAGATTAGAGTAAGGCTCATAAATTTGATTTTTAT	8683
Qy	4176	AAGGATATGGAGACAAAAATGAATGATTTAAATTTTCAGTTATTGTACCAATTTATTAATGT	4235
Db	8684	TTGATATGGTACTATTATAGAAATGCTTCTTAAGGTTTCTTAAAGTTTAAAGAAATAGTTATTAT	8743
Qy	4236	CCAAGATTATCTTGATAAATGCTATTACAGTATTATTAAACCAACATATACTAATTTTGA	4295
Db	8744	TGTAATAGAAGGACGAATGTTATATATTTTACTCCTCATTTTACAGAGGCTACCGA	8803
Qy	4296	GGTTATTTCTCGTAAATGATGGAAGTACTGATGTTCTCGAGAAAAATTTGCTTAAACTATAT	4355
Db	8804	TTATCCTATC--TGTATGNTTCTCTATGTAATCAGACAAATATAAAATTTTATATGGTTGAT	8862
Qy	4356	GAAGAACGATGGGAAGAAATTAATATTAACAAGAAAAATTAATGGCGGCTTAGACAGATGCTCG	4415
Db	8863	TGTTTGATGATGGTTCTCGAGGACAGTACTAAGGAATTTGTAAGTAATTTATATAAAAGAA	8922
Qy	4416	AAATTTTCGGACTAGAAACATGCCAGGTAAATATTTGCTTTTGTCCATTTCTGATGACTA	4475
Db	8923	TAAAGTTAGTATTCGTATCTGTA--TAAACGTAAATGGCGGTAAGCAITTCAGCCTTAA	8979
Qy	4476	TATAGAAGTTGCAATGTTTCGAGAGAATGTCATGATAATATAACTGAGTATTAATGCCGATAT	4535

[illegible]





```
ISKFPAPKAYOKATTGTADSOLEMTKTELELSFBEYLEMRDYALSISKGVETSTPDBEES
LEFLISTDMIEYIKIPSGEITNLPYLEKIGKQKQKVLSTGMAWHEEIHQAVNLRQNG
TTDIIISLHCTTEYPTTPYPSLNVLNHTLKDFKDLTIGSDHSIGSEVPAAAMGAE
VIEKFTLTDMEGPDKASATPDILAALVKGVRIVEQALGRFEKIPDPVEERKQIVA
RKSVALPKIKKGDIYSIENTVTRPGNGISPMNWDILQEAQDDFEDEVIRDSRF
ENOLPEL"
/translation="1094. .2494)
complement (1094. .2494)
/gene="cpsL"
/notes="synonym: SAG1162"
complement (1094. .2494)
/gene="cpsL"
/notes="identified by match to PFAM protein family HMM
PF01943"
/codon_start=1
/transl_table=11
/product="polysaccharide biosynthesis protein CpsL"
/protein_id="AA000044.1"
/db_xref="GI:22534191"
/translation="MNTKLLQSGFIYTLGNLLVQGLAFITLPIYTRVISTEVYQVS
LYVAMGIIITFGLQTSGLSSARVKYGEFKSYSGSAFSGVGNIFLIVILLIAPLFR
SFLAPLVGESIFELLMVQSYANVVTFFGQYFIQQOQSLANLILSLANAVSVALS
LFLTFHWSDDFLSVFGAFVPTIITGIVAFAYIYTHSKSYNPKYFPIVTVSIPILF
HLGHQNLNOLDRLMLARLYSTKEVAMYSFGSLGMIQIVLNSINMAWIPWFFDARK
EKLQLSTYSRYLGLVELTGLVTVPELQIOMGGDKYSSVQFSLIISYELVF
LYTFPVMIQFYANTWIPIGTLLAAGVNMLLNLVLIPIHYAAGAAMATIIISIALLI
FHHIVSKYKVSVDVRQYIIILSGIVFSYAMLMNMFIGNVIRWSLIGIILIVISIV
FQKVIDLLGKKRRR"
complement (2491. .3447)
/gene="cpsK"
/notes="synonym: SAG1163"
complement (2491. .3447)
/gene="cpsK"
/notes="similar to GP:13549137; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="polysaccharide biosynthesis protein CpsK(V)"
/protein_id="AA000045.1"
/db_xref="GI:22534192"
/translation="MNNFVCHTLVHLLITIKLKDKENTRIFCIDTITNYETVWKTIN
BOGRTRAVKEFAVELOQSKIIEVMELVDSLNLNHYPERVDVQVLPFNDOTLIGRYM
VYLGKHLLEDGNCQAKLFLGGSVVKVITKYLEKYPVGFPSKYLCSISVNSIV
GLPHDIRSKYKELPRKLFDSLNKEQSLIFKIFKTKPLTIYPKSVLLTQPLAQDK
CYKTPTRFQSIQDYFDIDIVQSYETLGNVYLVKVRDSDVDVSKLPVELLPSNIP
MEIIFELMTRGFEFGITHSSTALDFLTCDVKKTLVDLKD"
complement (3444. .4409)
/gene="cpsJ"
/notes="synonym: SAG1164"
complement (3444. .4409)
/gene="cpsJ"
/notes="identified by match to PFAM protein family HMM
PF00535"
/codon_start=1
/transl_table=11
/product="glycosyl transferase CpsJ(V)"
/protein_id="AA000046.1"
/db_xref="GI:22534193"
/translation="MDKVSIIIPYVNVQSFINECIESVLAQTSYSLNLIILVNDGSTDN
SGDTCDYSEIDGRIFVFKHNGKSDARNYGISRAGDYVILDSDDLYLKGDAIER
MVPEKYSNRIVLGCVKREQHIINIVLEDEMLETISPQATONLYNVDAIRAIFT
VAHNLKRELFSTLCYFVGKLEHDEFYTKYLKAKNIIIFRYNTAYAIRENSIMT
GSYNIKRHVAELKREYLLEKYPDLVQFOSERALIKMEVNLLIELYNNKPNFYKFFHTL
KTEYKKTIFDPKRRKMLKIKYLYKVCVHFPTLNCRRKKKINK"
complement (4402. .5385)
/gene="cpsO"
/notes="synonym: SAG1165"
complement (4402. .5385)
/gene="cpsO"
/notes="identified by match to PFAM protein family HMM
PF00535"
/codon_start=1
/transl_table=11
/product="glycosyl transferase CpsO(V)"
```

```
/protein_id="AA000047.1"
/db_xref="GI:22534194"
/translation="MELISIIIPVYNGEIVIGRCLDSILEQTYQNLEIILIDGSSDR
TGDICEKYLEDRRIKYFOENRQGSVARNNGVLRCTGDMWIAFLDSDVLPYSIEVM
YNIQKATNADIVLTSIGNFNNTYNTINSQYKEIKLYLLEVALEEMETGKTYGVSPF
AKLPKSNLLSNPYPGKTHEDMTTKLISQASKIACVDIVTAVVVSNDTRTKF
NERMLYFPAIQNNIVFINLPHNTSLISAVYNEVFGDIDICGKMWIDPKLYDVYD
YRKKYKFKTILFNRRISVKGKVKILFISSIRYFTIVRKIYNLRSLRNG"
complement (5389. .6276)
/gene="cpsN"
/notes="synonym: SAG1166"
complement (5389. .6276)
/gene="cpsN"
/notes="identified by match to PFAM protein family HMM
PF00535"
/codon_start=1
/transl_table=11
/product="glycosyl transferase CpsN(V)"
/protein_id="AA000048.1"
/db_xref="GI:22534195"
/translation="MLYIPTPTFNRGYRLSYLSDLCNQTNRNFIWLIVDDGSDSTK
EIVSNYIKENKVISIVLYKRNKGKSHAYNLAMRYMQPSDYHVCVDSDDMLLEDAVEII
FKDLESLTNSRVGLVPRYSLNQGNLNPKILEVNIPLDKYKHLKTCICVINN
AYLVDFEPCFEGENFLSEIMYIYLSKGYFCQPNRKLYCFDYLEDLTSTNIFKLWR
KNFKGTFISLNSYMYVMSFPNIFDRWNSAIKKNIQALKMTTLGVIPTLKSEBAGH
KILLGFSYLKWKARFKKSE"
complement (6302. .7027)
/gene="cpsM"
/notes="synonym: SAG1167"
complement (6302. .7027)
/gene="cpsM"
/notes="similar to GP:13549133; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="polysaccharide biosynthesis protein CpsM(V)"
/protein_id="AA000049.1"
/db_xref="GI:22534196"
/translation="MIPKVIHYCWFGNPLPDNLKKYIKTWBQCPDYEIEIWNHNY
DYSKNVMEAYTKNFAYVDYARDIITYYGGFYLDTDVLLKSLDPLRIHECFLA
RETSVDNGLIIGAVKGGHFLKSNMSIYDKSLTSLNKTCEVTVTNLLNRGLKNQK
LIQKIDITIPRYENPKNLLTKYKDLTSTVYSIHHEGSKSSSFISDSLKIRVR
LIIDFLFGYTRMLLRFLKKG"
complement (7066. .8160)
/gene="cpsH"
/notes="synonym: SAG1168"
complement (7066. .8160)
/gene="cpsH"
/notes="similar to GP:13549132; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="polysaccharide biosynthesis protein cpsH(V)"
/protein_id="AA000050.1"
/db_xref="GI:22534197"
/translation="MKINKNSLFYIAFLVNFPSLGLGEGNSAYKIVMLVAILLCGI
KFLDLSYFERKLVIFILLFIATILNLPFVHKVTIFLTLFFLAKDLISLKKAFSII
IGSRILGLNLQIFVKLDLIEIKYVNFYRDQGFILRSDLGFHPNTHNPFALTIFLY
IVLVYKRLKPVMLVFLTNLYLYQVTFSTGYIIVTLFVLYVTKNSLIRKFMKL
APVQFLLAVFTFLSTIPFNSNFVOKDLVLLNGLRHVAHLQVLDGTPFGNSFKETS
VFDNSYGLMSYGVVLFMFCMLIYIYKSKLIIILQLLFIIMSILITFESFPYSP
VMNISWLVGKIPCDGIEPIKKEFTIVNNI"
complement (8147. .8638)
/gene="cpsG"
/notes="synonym: SAG1169"
complement (8147. .8638)
/gene="cpsG"
/notes="similar to GP:5823214; identified by sequence
similarity; putative"
/codon_start=1
```

Query Match 11.4%; Score 798.8; DB 1; Length 21365;  
Best Local Similarity 48.2%; Pred. No. 2.5e-81;

	Matches 3323;	Conservative	0;	Mismatches 3382;	Indels 185;	Gaps 31;
Qy	16	TGCGAATTAATTGATATGATAGCAGTTGCGAATTTCTGCAATTTCTGCAATCTTAAACAAGTCATATACCAA	75			
Db	10465	TAGCGATGATTCACAAACAGTTGTGGTTTATTTTTCTGCAAGTTTGACATTTAAACATTAAATT	10406			
Qy	76	ATGCTGATTTAAATCG---TTCCTGGAAATTTTATCATTAATGATGGTTTCATTAATTTTGCAT	132			
Db	10405	CTCCCAATTTTAAAGCAATTAAGATTTATTTGTTCTTCTATTGATACATTAATATTGTCT	10346			
Qy	133	TTTTTATATCTCTGATGCCAGTTGAAATTTTGAGTATAGAGTAAATCTGATAGAGTTTGGAAA	192			
Db	10345	TTTTATCTTCTGATTTTACAGAGACTTTTGGAGTCGTGGCTATCTTGAAGAGTTTAAAA	10286			
Qy	193	AAACATTTAACTATAGTATAATATTTTGCAATTTTCTTACGGCAGTATCATTTTGTGG	252			
Db	10285	TGGTATTGAAATACAGCTTTTACTATATTTTCATATCAAGTTCAATTATTTTTTATTTT	10236			
Qy	253	AGAATAATTTTCGCACTTTCAAGACGTGGTGGCCGTGATTTTTCACATTAATAAATCTTCGTTT	312			
Db	10225	AAAACTCTTTTACAAACGACACGACTTTTCCCTTTTATCTTTTATTTGCTATGAAATTCGATTT	10166			
Qy	313	TGCTATACCTATTTTAAAGTAATTAATTAAGCAGTTTAAAGGATAGCTTTCTATTTTTCGACAA	372			
Db	10165	TATTATATCTATTGAAATTCATTTTAAATATTATTCGAAAAATTCTTTACGCTAAAGTTT	10106			
Qy	373	TCTATCAAAAAAGACGATTTCTAATTAACAACGGCTGAACGATGGGAAAAATATGCAAGTTT	432			
Db	10105	CACGAGATACCAAGTTGTTTGATTAACGATAAGGATTTCTTTATCAAAAAATGACCTTTA	10046			
Qy	433	TATTTGAAATCACATAAACAATTCAAAAAACTCTGTTGCAATTTGCTGATGGTTTGTAGTACAG	492			
Db	10045	GGAAATAAT---ACGACCATAATATATCGCTCTGCTATCTTGGATTCCTCTGAAAAAGG	9989			
Qy	493	AAATAGATAAAATTAATTTATCATTAACCGCTCTATTATTCTGTGGAGAGCTCTATAGAGT	552			
Db	9988	ATTGTTATGATTTGAAACATPAACTCGTTAAGGATAATAAACAAGATGCTCTTACTTCAG	9929			
Qy	553	TTTCAACAAGGGAGTGGTCCACACGCTCTTTATAAATCTACCAAGTGAGTTTGTAGACG	612			
Db	9928	AGTTAACCTGCTTAACTGTTGATCAAGCTTTTATTAACATACCCATTGAAATTTTGTGTA	9869			
Qy	613	TAAAGCAATTC-----GTTTCAGATTTTGAGTTGTTAGTATPTGATGAAGCGTTG	663			
Db	9868	AATACCAAAATACAAGATATTTAATTAATGACATTTCAAGCAATGGGAGTGAATGTCAATGTTA	9809			
Qy	664	ATATTAAATTCATTCGGTTTTACTCGGTTGAAAAACAACAAAAATCCAACTGCTAGGTGACC	723			
Db	9808	ATGTAGAGGCACCTTAGCTTTGATTAATATAGGAGAAAAGCGAATCCAAACTTTTGAAGGAT	9749			
Qy	724	ATAGCATGTGAACTTTTTCACAAATTTTATAAGCCTAGTCAATATCATGATGAACACGAC	783			
Db	9748	ATAGTGTATTACATATTCTATCGAAATTTCTAATAATATAGTCACTTTATGCAAAACGAT	9689			
Qy	784	TTTTGGATATACTCGGAGCGGTAGTCGGGTAAATTAATTTTGTGATAGTTTCTTATTTGT	843			
Db	9688	TTTTTGGATATCACGGGTGCTATTATATAGTTTGCTCATATGTGGCAATTTGTGGCAA	9629			
Qy	844	TAGTTCCAATTAATTCGTAGAGATGTGGACCGGCTATTTTGTCTAGAAAACGAGTTGGAC	903			
Db	9628	TAGTTCCACAAATCAGAAAAAGATGGTGGACCGGCTATCTTTCTCAAAATAGAGTAGGTC	9569			
Qy	904	AGAAATGACGCATATTTACATTTCTACAAGTTTTCGATCGATGATGTTGATGCTGAGGAGC	963			
Db	9568	GTAATGTGAGGATTTTATAGATTTCTATAAATTCAGATCAATTCGAGTAGATGACAGACAA	9509			
Qy	964	GCAAAAAGACTTGTCTCAGCCAAAAACAGATGCAAGGGTGGGTATGTTTAAAAATGGGAA	1023			
Db	9508	TTAAGAAGATTTATTAGTTTCAACATCAATGACAGG-----CTAATGTTTAAAGTTAGA	9454			
Qy	1024	AAACGATCTTAGAATTAATCTCAATTTGCAATTTTCATACGCAAAAAACAAGTTTAGACGAGT	1083			
Db	9453	CGATGATCTCTAGAAATTAATAAATAGAAAAATTTATTCG-AAAAACAACATAGATGAGT	9395			

QY	2147	AGAAGAAATTTTCAAGATAATAATATTTTATTTATAGAAAAATATAGATGATTTGTTTGGAA	2206
Db	8314	AATAAGGTAAAAACAATGTGTAAATTTTGTATATCGTTAGATATTTGAAGGTTTACAAAT	8255
QY	2207	AAAATTTTGAAGTTTCTAAGCAAACTAACTTTTACATCAATATATTTT- - -TTTGTG	2263
Db	8254	GTAGTCTATGAGGGAATGATGAATCGTCCGTTTTTGAACCTATATAGTAGTAATTTTAT	8195
QY	2264	GAAGAATTAAAACAAATAGTTGAAAAATTTAAATGAGGATCAAGAAAAATGAATAAAAA	2323
Db	8194	GAAGAAATTTAAGGTAAATATTAAAGGAGTGTGCGATGAAAAATCAATAAACTCTTTAT	8135
QY	2324	AGATGCATATTTGATAATAGGCTTATCAATAATTTTTTCTCAGATTTTACTGAGAGGGATAC	2383
Db	8134	TTATATTGCAATATTTTGTAGTTAATTTTTTAAATCACTGGGTTTAGCGGAGGGAACTC	8075
QY	2384	AGATATTATCATCTTCTCTCAGGAGATGCACCAATTAGTTCCTTCAGATACCTGTAT	2443
Db	8074	AGCTTACAAATAGTG-----ATGTTAGTTGCAA	8046
QY	2444	AAATATTTTAAATATCTCAGGATTTATATGTTTGAAATTTACAAAAGATGAGCAAAATAT	2503
Db	8045	TTTTACTGTGTGGAAATAAAATTTTTTATAGTAGCCCTTATTTTGAAGAAGAAACTCG	7986
QY	2504	AAAGAAAATAGGATATATGAACGAGTTAAATGTTTACAGATTTATTCCTAATATACGAA	2563
Db	7985	TGATCATCTTTTATTTATTTATCGGACCAATTTTGAATTTATCT-----	7941
QY	2564	AAACTATTGATATGACTGTTTGAATTTTTTATTAAGATGTATCGAGCTTTTGAATAC	2623
Db	7940	----TTGTTCTATAAGGTTACTTTTATATATAACTTTAAATTTTTTCTCAGCATTTAAAGGAC	7885
QY	2624	TATTTACAAAGATTTGTTTATTGATAGATAAAAAACATGGTCTCAAGAAATAAGATTTC	2683
Db	7884	ATCTCTCTAAAAAGCTTCTCTATATATATAGGATCGCGTATTTTGGGAGTCTATTAT	7825
QY	2684	GTTCTAATTTGGGTTTCGCTCCACATGATTTTGTGGCAATTTCTTTATCAAAATGAAAACG	2743
Db	7824	AATCAAATTTTGTGAAATTAGATTTAATAGAAATTAAGTATGTCAATTTTTTATAGGAT	7765
QY	2744	AAACAGCTTATTATTTTAAAGTAACTAAATGTCAGATGAACATTTATTATACAGCAATTA	2803
Db	7764	GGACAAATTTATCTTGAGAAGTGA--CTTAGGTTTGTGTCATCCTAACTTTATTTCATAATTT	7706
QY	2804	TAGAAAATATGAATTTTCAAAATAGATTCTAAATATGGAA-- -ATTTAAGATATATAA	2860
Db	7705	TTTTGCTCTACTATTTTCTTGATATTTGTACTCAATATTAAGAGCTAAAGCCGTGTGT	7646
QY	2861	AGTGGAAAAATCAACATCTTCTCCTATTTGTTTACAGATGATTTCTATTGATGAATTCG	2920
Db	7645	GATGGTTTTATTTTTTAACATTAATTTATTTGTCACCAATATACTTTTTCAAGGACAG	7586
QY	2921	TAAATGCAAGAAATTTAGGTTTTTTTATTTGCTAGAAAATTAATAAGAAAAATAAATCTA	2980
Db	7585	GTATTATATCGTAATTTTTTATTATTGTACTCATTTATGTCACAAAGAAATAGCTTAATAA	7526
QY	2981	AA---- -TTTAAAGAAATTTTACTTAAAAAATAAATAGTTTCTTTCTGAGAGTAAATGA	3036
Db	7525	AAGAGTATTATGAAATTTAGACCCTATGTACAAATTTTTTTTATAGTATTTTACCTTTTT	7466
QY	3037	TGTTTTAAATTTTAAAAATATGACCCGGAATTTTTTATTTTTTAAGTACTTCTGGTTGATTA	3096
Db	7465	GAGTTCTACAATTTTTTTTAAATTAATTCAAATTTGTTTCAAAATTAGATGTTCTTTTACAGG	7406
QY	3097	TTTTTATTTCCAGACAAAAGATGATATTTTTATTAATTTTTTATAGAAATTTAATTTTATTC	3156
Db	7405	TAGATTACACTATGCTCATTTTACAACCTGTAGATGGTTAACTCTTTTTCGAAAATAGTTT	7346
QY	3157	ATATAAAAATTTTGAAAACCTAAGCTAATTTTAAAAAATGAAATTTTATTTGTTTTATTAT	3216
Db	7345	TAAGGAACCAAGTCTCTATTGTAATATAGTACTCTATGTTTATGTAGTATGTATGTGT	7286
QY	3217	GGTCTATATTATGTTTTGTTTTTTCAGTAGTCAAGATGTTTTGTTTGAATAAATTTTGGAA	3276

[illegible]

Db 6234 TTATCCTATC-TGTAAGATTCCTAAGTAATCAGACAAATAAAATTTTATATGTTGAT 6176  
Qy 4356 GAAGAACGATGGAAGAAATTAATATTAACAAGAAATTAATGCGCTAGCAGATGCTCG 4415  
Db 6175 TGTTGATGATGTTCTGAGGACAGTACTAAGGAATTCGTAAGTAATTAATAAGAGAA 6116  
Qy 4416 AAATTTCCGACTAGAACATGCAACAGGTAAATATATGCTTTTGTGCGATCTGATGACTA 4475  
Db 6115 TAAAGTTAGTATGTTATCTGTGTA---TAAACGTAATGCGGTAAAGCAATTCAGCCATAA 6059  
Qy 4476 TATGAAGTTCGAATGTTCCGAGAGATCGATGATTAATACTGAGTATTAATCCGGATAT 4535  
Db 6058 TCTGCAATGAGATATATGCAACCAAGTATCATGTAATGTTAGATAGTATGATG 5999  
Qy 4536 AGCAGATAGATTTTGTGTTAGTAGACGAAACCGGTATACAAAGAAAAAGAAATAG 4595  
Db 5998 GTTATTAGAGATGCAATGTAATTTTAAAGATTTAGAGCCCTAAAGTTCTCTAA 5939  
Qy 4596 TAAATTTTCATGCTTTAAACGAGAGAGACCTGTAAAGAAATTTTGTGAGATCTAATAT 4655  
Db 5938 TAGATATGAGCCCT-----TGTTTATCCTCGATATTCATTAACCAAGGTAAATAATG 5885  
Qy 4656 AGAAATATGTTGGTCCAGCTTTATTCAGGATATTAATAAGATATAAAATTTCCA 4715  
Db 5884 GTTGAATCCCAAAATATTAGAAATTAATTCCTGATTTAAATATAAATATCACTTAA 5825  
Qy 4716 AATTAAATATAGAAATATTGGTGAGATTTGCTTTTAAATTTGGAGGTCTTGAACAAATG 4775  
Db 5824 AATTGAACTTGTATTGTTATTAATAA--TGCCTATTAGTAGATTCGAATTTCTTGT 5767  
Qy 4776 AACAGCTGATGATGTAATCAGAGAAATATTAATTAATTAATGATTCGTAAACAGTTC 4835  
Db 5766 TTCGAAGGGGAGAAATTTCTTATCAGAGAAATAATGATATATATTTATCAAAAGGGA 5707  
Qy 4836 GCTTATTAATCAGAAATTTCTCTATAAATAATTAATGATTCAGACAAATTTGGAGAAATTA 4895  
Db 5706 TACTTTTGTCCCAAAATAGGAAATCTATTG---TTTTCGACTACTTAGAAGATGGTTT 5651  
Qy 4896 CCCCTTTAAGTTAAAGAGAGATTTAGTCAATTTATTTGATGCAAAAGTTATTAAGAGAA 4955  
Db 5650 GACTTCTAATATTTTAAATTTATGGAGAAATAATTTCAAGGACTATTTTTCATTAGA 5591  
Qy 4956 GGTAAATGTTTAAACAAATGTAATCAACAGATTTGTTGGATATGATGTTCTTGCCAAAT 5015  
Db 5590 AAATTCATATATGATGTAATGCTCTTTCTTCAATATATTTGATCGCTGCTGCTCAGC--T 5533  
Qy 5016 ATTAGAGTCTTATCGAAAGAAATACGATAGATATCCATTTATTAAGCGAAAGAAATATTT 5075  
Db 5532 ATAAAGATAAAATGAACATACAGGCTTTGAAATGACTACTTTAGGGGTAAATACCTACA 5473  
Qy 5076 ATCAAGAAAGCATTTAGTGTGTTGTAATTTGATGAAATTTTCGCTAACTATATGTAAT 5135  
Db 5472 CTAAG-----AGTGAAGAGCTGGGTGGAAAAATTTTGTGGGTTTGTAGCTATCTT 5422  
Qy 5136 GTTATATAAGAAATTTCAAAAGCAGTAGAGGTAAAAATGGATAAATTTAGTGTATTGTT 5195  
Db 5421 TGGAGGTAGCAAGTTTAAAGAGCCAGTAAGAAATGGAAATTAATTTCAATAATAGTG 5362  
Qy 5196 CCAAGTTTATAATGTAATAATTTAATAGTATGTTGTATAGAAAGCAATTTAATCAAAAT 5255  
Db 5361 CCGTATATAATGGAGAAATATATATGGAGATGCTAGATAGTATTTCTTGGACAAACT 5302  
Qy 5256 TATAAATATAGAAATATTTGATAGATGATGCTCTGTAGATGATTTCTGCTAAATA 5315  
Db 5301 TATCAGAAATTTAGAGATTTATCAATAATGATGATGCTCTAGTGAATCGCACAGGGGATAT 5242  
Qy 5316 TGCAGGAATATGCAAAAGATATAAGAGATAAAAATTTTTTCACTAATCATAGTGA 5375  
Db 5241 TCGGAAAGATATTTTGGAGAGACAGGCGAATAAATAATTTCTATCAAGAAAAACAGAGC 5182  
Qy 5376 GTATCAAAATGCTAGAAATCATGGAAATAAGCGGAGTACAGCTGGAATATATTTGTTT 5435  
Db 5181 CAATCAGTTGCAACGTAACCAACGAGTACTTAGTGTACAGGAGATTTGGATTTCTTTTA 5122

Qy 5436 GACTCTGATGATGTTGTTGATAGTAGATTAGTAGAAAAATTTATATTTTAAATTTATATAAA 5495  
Db 5121 GACAGTGATGATGTTTATCTCTCTTATTCATTTAGAGTTATGTAACAACATACAGAAAGCT 5062  
Qy 5496 AGTAGAGTATGATTTATCTGGTGTGTTGTAACCTACTTTTTCAGAAAAATATAAATAATTTT 5555  
Db 5061 ACCAATCCGATATTTGTTTAACTAGTATTTGTTAACTTCAACAATCTTCAACTACTAGT 5002  
Qy 5556 GAAGTGAATATCCAAATATTGATTTTGAAGCAATTAATACCGTCGACGACATGCGAGAA 5615  
Db 5001 ATTAAATTCACAGTATTTAAAGAGATTAACCTTTTACTTTTAGAGGTGCGCTTGGAGAA 4942  
Qy 5616 AAAAATTTTATGAATTTGTAATATAAATAATTTTCTACTCTCTGTTTGTAAACTATAT 5675  
Db 4941 ATGTATTATGGAATACTTATGGAGTTTCCCGCTTAGCAA-----AATTATATCCGAGAA 4887  
Qy 5676 AAGAAAAGATACATAACAGATCTTTTTCAGAGAAATCAATCGTTAGGAGAAAGATTTACTT 5735  
Db 4886 GTAAATTTTATGAGTAATCCATATCCAGAAAGGAAAAATTCATGAAGATATGGATACAACCT 4827  
Qy 5736 TTTAATCTGCAATTTTAAAGAAATATAGATAGATTTAGTATTTTGAATTTGACATGACATCTTAT 5795  
Db 4826 TTAAGCTTATATCTTGGCTTCAAAAATAGCAGTTTGTGATATTTGCTACTCGAGTGTGT 4767  
Qy 5796 TTTTATAGGAGAGGTATATAAGTACAGTAAATTTCTTTTAAAGAGAGGTGTGTTTTGCAA 5855  
Db 4766 ATTTTAGCGATAATAGTACTACTAGGACAAAATTCATAGAGAAATGT-TGTAATTTTTT 4708  
Qy 5856 TTGGAATTTTGGAAAAACAAAGTATGATTTGTTTAAAGAAATATATATGTTAGGAGATTTT 5915  
Db 4707 GAGCAATTTCAAAATAATATTTGTAATTTATAATCTAAATTTTCCACATAATACATCTTTA 4648  
Qy 5916 GACGTATCAATTTGTTAAAGATATATAGTGTGCAAGTATTTTATATATAGCTTACTAATG 5975  
Db 4647 ATCTCTGCTGTAATATAATGAAGTATTTGCTGGAATAGATATATGTTGGAGAGATTT 4588  
Qy 5976 TTTAAATACGGAACACAGTCTATTTTTCACAAAATTTTAAATTTTGTAGAAATCTTTTATAA 6035  
Db 4587 GATTTTAAATATATATGATACAGTGTGATTTATACAGAAAAAATATAGAAAAATTTTAAAG 4528  
Qy 6036 AAATATTTTAACTGTTTAAAGATATCTAACAAAAATCTTTGCTTAAAAATTTTGT 6095  
Db 4527 ACTATCTCTTTAATAATAGGATTTAGTGTGAAGGAGAGGTTAAATATATATTTATTTAT 4468  
Qy 6096 ATAGAAATTTGTTGCAACAAAGTTTAAAGAAATATTTATGTTTATATAGAGATATC 6155  
Db 4467 TCCTCTATAAGATATTTCACTATTTGTAAAGAAAAATA-----TATAATTTGAGGTATC 4415  
Qy 6156 ATGATATCTATTAGTAAAAATTTCTAATTTGTACTATATATATATATAGTAAAGAAATATTTA 6215  
Db 4414 AAGGA-----ATGGATAAAGTTAGTATCATTTATCCAGTTTATACGTTCAATCTTCTCTA 4359  
Qy 6216 TCTAAATGTATAGATAGCATTTGTAATCAGACCTACAAACATATATAGAGATTTCTCTGGTG 6275  
Db 4358 AACGAGTATAGAAAAGTGTGTTTGTAGCTCAAACTTATTCAAATTTTAAAGAAATTTCTAGTT 4299  
Qy 6276 AATGACGATGATACGATTAATTCGAGAAATTTGTTTGTAGCATATCGGAGAAAGATAGT 6335  
Db 4298 AACGATGGCTACAGATAATTCGCGGATATCTGTGATTTATTTTCAAGAAATAGATGGG 4239  
Qy 6336 CGCAATCTGTTATTTTAAAAAGAGAACGCGGGCTATCAGATGCGGTAATTTATGCGCAT 6395  
Db 4238 CGTATATTCGTTTTCACAAAAATATGTTGGTGTGTTTATCAGATGCTCGTAATTTATGAAAT 4179  
Qy 6396 AGTCGCGCAAGGTGACTACTTAGCTTTTATAGACTCAGATGATTTTATTTTATTCTCGGAG 6455  
Db 4178 AGTAGAGCTACTGGGATTTATTTTATTTACTTGTATTCGGATGATTTCTATATAAAGAA 4119  
Qy 6456 TT---CATCCAACTTTACAGAAAGCAATTTGAGAGAGAAATGCCCTTGTGGCAGTTGCT 6512  
Db 4118 GATCGAATCCGAGAAATGTTTGAATTTTAGCGGAAATAATAATTCAGAAATTTGTTTAGGG 4059

Qy 6513 GGTATG-----ATAGGCTAGATGCTCGGGGCAATTTCTTAACAGCAGAGCCGCTTCCT 6566  
|||||  
Db 4058 TGTATGTTGAAAGAGAGACACATATTATAATATCGTTTATAGAGATGAAATGATA 3999  
Qy 6567 ACAATCAGGCTGTTCTGAGCGGAGGAATGTTTGTAAAGCTGCTAGAGCGGATGGT 6626  
|||||  
Db 3998 GAAACTATTAGTCCAGTACAGCAATACAAATATTTTATAA---TTATGATGCTATAGA 3942  
Qy 6627 CATCGCTTTGGTGGCTGTATAAATCTCTATAAAGAGACTATTTTGAAGATTTTCGA 6686  
|||||  
Db 3941 GCTATTTTACTGTAGCACATAATAAATATATAAGAGAAATTTTCTACGCTGTC 3882  
Qy 6687 TTTGAAAGGTAAGATTCATGAAGATGAATCTTCACTTATCGCTTGTCTATAGATTA 6746  
|||||  
Db 3881 TATCCAGTAGGAAGTTACAGAGATGAATTTTAACTTATAAATTTATTTTGAAGCC 3822  
Qy 6747 GAAAGGTGCAATAGTTAAGGAGTCTTGTACTATTTATGTTGACCGAGAAATAGTATC 6806  
|||||  
Db 3821 AAAAATATAATTTCTTTTAGATAATAACTTATGCATATCGTATTAGAGAAATAGCAT 3762  
Qy 6807 ACAACTTCTAGCATGACTGACCATCGCTTCCATTTGCCCTACTGGAAATTTCA 6856  
|||||  
Db 3761 ATGACTGGTCTCTACAATATTTAAAGGCTACATGCTGTAGAAGCACTTAA 3712

RESULT 10  
LOCUS CQ655070 349980 bp DNA linear PAT 02-FEB-2004  
DEFINITION Sequence 12027 from Patent WO234771.  
ACCESSION CQ655070  
VERSION CQ655070.1 GI:41687947  
KEYWORDS Streptococcus agalactiae  
SOURCE Streptococcus agalactiae  
ORGANISM Streptococcus agalactiae  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.

REFERENCE 1  
AUTHORS Telford, J., Masignani, V., Margarit y Ros, I., Grandi, G., Fraser, C.  
and Tettelin, H.  
TITLE Nucleic acids and proteins from streptococcus groups a & b  
JOURNAL Patent: WO 0234771-A 12027 02-MAY-2002;  
Chiron S.p.A. (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)  
FEATURES  
Location/Qualifiers  
1. 349980  
/organism="Streptococcus agalactiae"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:1311"  
/note="seq 10967 to big: splitted in-seq 10.967, from  
0.000.001 to 0.349.980-seq 12.025, from 0.300.001 to  
0.649.980-seq 12.026, from 0.600.001 to 0.949.980-seq  
12.027, from 0.900.001 to 1.249.980-seq 12.028, from  
1.200.001 to 1.549.980-seq 12.029, from 1.500.001 to  
1.849.980-seq 12.030, from 1.800.001 to 2.149.980-seq  
12.031, from 2.100.001 to 2.160.266"

ORIGIN

Query Match 11.4%; Score 798.8; DB 6; Length 349980;  
Best Local Similarity 48.2%; Pred. No. 9.7e-82;  
Matches 3323; Conservative 0; Mismatches 3382; Indels 185; Gaps 31;

Qy 16 TGGCAATATTGATGATAGTAGCAGTTGCAATTTCTGCAATCTTAAACAAGTCATATACCA 75  
Db 247235 TAGCGATGATTCAAACAGTTGTGGTTTATTTTCTGCAAGTTTGCATTAACATTAATTA 247176  
Qy 76 ATGCTGATTTAAATCG---TTCTGGAATTTTATCATAATGATGGTTTCATTTTTCAT 132  
Db 247175 CTCCCAATTTTAAAGCAATAAAGATTATTGTTTGTCTTCTATGATCAATTTATTTGCT 247116  
Qy 133 TTTTATATCTCGTATGCCAGTTGAAATTTAGTATAGAGGTAATCTGATAGAGTTTGA 192  
Db 247115 TTTATCTTCTGATTTTACAGAGACTTTTGGAGTCTGGGCTATCTTTGAAGAGTTTAA 247056  
Qy 193 AAACATTTAACTAGTATATAATTTTGAATTTTCTTACGGCAGTATCATTTTGTGTG 252

Db 247055 TGGTATTGAAATACAGCTTTTACTATATTTTTCATATCAAGTTTCATTTATTTTATTTT 246996  
Qy 253 AGAATTAATTTTCGACACTTTCAAGACGTGGTGGCGGTATTTTCATTAATAATAAATTCGTTT 312  
Db 246995 AAAAATCTTTTACACGACACGACTTTTCCCTTTTATTTTATTTGCTATGAATTCGAATTT 246936  
Qy 313 TGGTATACCTTATTAAAGTAATTTAAAGCAGTTTAAAGGATAGCTTTCTATTTTCGACAA 372  
Db 246935 TATTATATCTATTGAAATTCATTTTAAATATTTATCGAAATATTTCTTACGCTAAGTTT 246876  
Qy 373 TCTATCAAAAAAGAGAGATTTCTAATTTACACGGCTGAACGATGGGAAAAATATGCAAGTTT 432  
Db 246875 CACGAGATACCAAAGTTGTTTGTATAACGAATAAGGATTTCTTTATCAAAATGACCTTTA 246816  
Qy 433 TATTTGAATCACATAAACAATAATTTCAAAAAATCTTTGTTGCAATTTGGTAGTTTATAGGTACAG 492  
Db 246815 GGAATAAAT---ACGACCATAATTTATATCGCTCTGTATCTTTGGATTTCTCTGAAAGG 246759  
Qy 493 AAATAGATAAAAATTAATTTTATCATTTACCGCTCTATTATTCTGTGGAGAGAGCTATAGAT 552  
Db 246758 ATTGTTATGATTTGAAACATAACTCGTTAAGGATAATAAACAAGATGCTCTTACTTCAG 246699  
Qy 553 TTTCAACAAGGGAAGTGGTCGACCACTCTTTATAAATCTTACCAAGTGAAGTTTATAGACG 612  
Db 246698 AGTTAACTGCTTAACTGTTGATCAAGCTTTTATTAACATACCACTTGAATTTATTTGGTA 246639  
Qy 613 TAAAGCAATTC-----GTTTCAGATTTTTCAGTTTGTAGGTATTTAGTATGATGAAGCGTTG 663  
Db 246638 AATACCAAAATACAGATATTTATTAATGACATTTTGAAGCAATTTGGGAGTGAATTTGCAATGTTA 246579  
Qy 664 ATATTAAATCATTTCCGTTTACTGCGTTGAAACAACAAAAATCCAACTGCTAGGTGAC 723  
Db 246578 ATGTAGAGGCATTTAGCTTTGATAATATAGGAGAAAGCAATTTTGAAGAT 246519  
Qy 724 ATAGCATTTGTAATTTTCCACAAAATTTTATAAGCCTAGTCTATATCATGATGAAACGAC 783  
Db 246518 ATAGTGTATTACATATTTCTAGAAATCTTATAAATATAGTACCCTTATAGCAAAACGAT 246459  
Qy 784 TTTTGGATATACCTCGGAGCGGTAGTCGGGTTAAATTTATTTGTTGTAATGTTTCTATTTTGT 843  
Db 246458 TTTTGGATATACCGGGTGTCTATTATAGTTTGTCTATGTCATGTCATGTCGCAATTTTTC 246399  
Qy 844 TAGTTCCTCAATTTTCTGTAGAGATGTTGACCGGTATTTTGTCTCAGAAACGAGTTGAC 903  
Db 246398 TAGTTCCTCAAAATCAGAAAGATGGTGGACCGGTATCTTTCTCAAAATAGATAGTCT 246339  
Qy 904 AGAATGACGCATATTTTACATTTCTCAAGTTTTCGATCGATGTATGTTGATGCTGAGGAGC 963  
Db 246338 GTAATGTTAGATTTTGTAGATTTCTATAAATTCAGATCAATGCGAGTAGATGCAACAAA 246279  
Qy 964 GCAAAAAGACTTGTCTCAGCCAAAACAGATGCAAGGTGGGTATGTTTAAATGCGAA 1023  
Db 246278 TTAAGAAAGATTTTATTTAGTTTCACAATCAATGACAGG-----CTAATGTTTAAAGTTAGA 246224  
Qy 1024 AAACGATCTAGAAATTAATTTGACATTTTTCAGCAAAACAAAGTTTACAGAGT 1083  
Db 246223 CGATGATCCTAGAAATTAATTTTAAATAGGAAATTTTATTCG-AAAAACAGCATAGATAGT 246165  
Qy 1084 TACCACAGTTTATTAATTTTAAATTTGCGGATATGATCTAGTTGGTACAGCTCCACCTA 1143  
Db 246164 TGCTCAATTTCTATAATGTTTAAAGGTGATATGATGTTTAGTAGGAACACGCCCTCCCA 246105  
Qy 1144 CAGTTGATGAATTTGAAAAATATCTCTCGGTCAAAAGACAGATGAGTTTAAACAG 1203  
Db 246104 CAGTTGATGAATTTGAAAAATTAATTTCAACGCAAGACGACGCTTTAGTTTAAAGCAG 246045  
Qy 1204 GGAATACAGTCTCTGCGAGTTAGTGGTGTAGTATATATACAGACTTTCGACGACGTAG 1263  
Db 246044 GAATCACTGTTTGTGGCAAAATCTGTTGTAAGAAATAATATTTACTGATTTTGTGAAATCG 245985  
Qy 1264 TTCGTTTGGACTTTAGCATACATTTGATAAATTTGGACTATCTGGTTCAGATATTTAAATTTTAT 1323  
|||||

Db 245984 TAAAGTTAGATGTTCAATATATCAATGAATGGTCTATTTGGTCAGATATTAAGATTATTC 245925  
Qy 1324 TAAAGACAGTGAAGTGTGATTTGTTGAGAGGGAAGTAAGTAAAGTAT 1373  
Db 245924 TCCTAAACATTAAGGTAGTTTACTCGGGAAGGAGCTAAGTAAAGGTAAAGTTTGAAG 245865  
Qy 1374 -----ATGAAAGTTTGGTGGTGGTCTTCAAGGGGGAATTTGACTCACTTGTATTT 1426  
Db 245864 GAATATATGAATAATTTCTGCTGGTGGTCAAGTGGTGGTCACTAGCACACTTTGAACCT 245805  
Qy 1427 GTTAAACCGTTTGGGAAGGAAGAACGTTTGGGTAAACATTTGATATAAGAGGATGC 1486  
Db 245804 TTTGAACCCCAATTTGGGAAAAGAGATAGGTTTTGGGTAACTTTGATAAGAGATGC 245745  
Qy 1487 AAGAAGTCTTTTGAAGAAATGAATAATGTATCCATGTTACTTTCCAAACAAATCGCAATCT 1546  
Db 245744 TAGGATATCTAAGAGAAGATGTATATCAATCTTCTTTCCAAACAAACCGTAAGT 245685  
Qy 1547 CATTAATTTAGTGAATAATCTTCTTAGCTTTTCAAAATTTTACGTGATGAGAAACGAGA 1606  
Db 245684 CAAAAACCTTGGTAAAAAATCTATTTCTAGCTTTTAAAGTCTTTAGAAAAGAAAGACACAGA 245625  
Qy 1607 TGTTATATTTCACTGCTGGCGGCTTGTCTGCTCCCTTTCTTTTACATCGGAAACATTT 1666  
Db 245624 TGTTATCATATCATCTGCTGGCGCTGTAGCAGTACCATTTCTTTATATTTGGTAAGTTAT 245565  
Qy 1667 TGGAGCAAGACGATTTATTTATTTGAAGTATTTGATCGAGTTAATAATCTACATTAACCTGG 1726  
Db 245564 TGGTTTGAAGACCGTTTACATAGAGTTTTCGACAGGATGATATAACCACTTTGACAGG 245505  
Qy 1727 AAAAAGTATTTCCCGTAAACAGATATTTTATTTGTTCCAGTGGGAAGAAATGAAGAAGGT 1786  
Db 245504 AAAATTAGTATCTCTGTAACAGATAAATTTATTTGTTCCAGTGGGAAGAAATGAAGAAGGT 245445  
Qy 1787 ATATCCTAAATCTATTAATCTTGGGAGTATTTTATTAAGTATTTTGAACAGTAGGAAT 1846  
Db 245444 TTATCCTAAGCAATTAATTTAGGAGGAATTTTAAATGATTTTGTCCACATGGGGACA 245385  
Qy 1847 CATGAACCAACAGTTTAAATCGATTGATAAAGAGATTTGATTTATGAAAAAATGGAAGT 1906  
Db 245384 CATGAACAGCAGTTCCACCGCTTTATTAAGAGTTTGAATTAAGAGGACAGGTGCT 245325  
Qy 1907 ATAACCGACGAATAATTTATTTCAACAGGATATTTCTGACTATTTTCCAGATAATTTGCAAG 1966  
Db 245324 ATTGATCAAGAAGTGTCTCATTTCAACGGGTACTCAGACTTTGAACTTCAAGATTTGTCAG 245265  
Qy 1967 TATAAAAATTTCTCAGTTTACAAGAAATGGAACAAATATTTAAACAAATCAGAGTAGTT 2026  
Db 245264 TGGTCAAAATTTCTCATATGATGATATGAATCTTACATGAAGAAGCTGAGATTGTT 245205  
Qy 2027 ATTTGCCAGGAGGCCCGCTACTTTTATGAATTCATTTATCCAAAGGAAGAAACAAATTA 2086  
Db 245204 ATCAACATGGCGGCCCGAGCGTTTATGAATGCAAGTTCTTAAAGGAAAAAACAATAT 245145  
Qy 2087 TTGTTTCTTAGACAAAAAAGTATGGTGAACATGTAATGATCATCAAGTAGAGTTTGTGA 2146  
Db 245144 GTGGTCTTAGACAGAAGACAGTTTGGAGAGCATGTGAATATCATCAGTGCAGTTTGT 245085  
Qy 2147 AGAAGAAATTTTACAAGATAATAATTTTATTTATAGAAAAATATAGATGATTTGTTGAA 2206  
Db 245084 AATAAGGTAAAAACAATGTATAATTTTGTATATCGTTGTAGATATTTGAAGGTTTCAAAAAT 245025  
Qy 2207 AAAATTATGAGTTTCTTAAGCAACTAATTTTACATCAATTAATAATTTT---TTTTGT 2263  
Db 245024 GTAGTCTATAGGGAATGATGAATCGTCCGTTTTTATGAACATTAATAGTAGTAATTTTAT 244965  
Qy 2264 GAAAGATTAAACAAATAGTTTGAATAATTTAATGAGGATCAAGAAAAATGAATAATAAAAA 2323  
Db 244964 GAAGAAATTAAGGTAATTAATAAGAGATTGCGATGATAAATCAATAAAAACTCTTTAT 244905  
Qy 2324 AGATGCATATTTGATAAATGGCTTATCATAATTTTCTCAGATTTTCTCGAGAGGGATAC 2383  
Db 244904 TTATATTGCAATATTTTATAGTTAAATTTTAAATCACTGGGTTTAGCGGAGGAAACTC 244845

Qy 2384 AGATATTATCATCTTCTCTCAGGAGAAATGCACACCAATTAGTTTCTCTCAGATACTGTAT 2443  
Db 244844 AGCTTACAAAATAGTG-----ATGTTAGTTGCAA 244816  
Qy 2444 AATTATTTTAAATATTCTCAGGATTTATATGTTGAATTTTACAAAAGATGACGCAAAATAT 2503  
Db 244815 TTTTACTGTGTGGAATAAANAATTTTATTTAGATAGCCCTTTATTTTGAAGAAGAAACATCG 244756  
Qy 2504 AAAGAAAATAGGATATATGAACGAGTTTAAATGTTTACAGATTATTTCCCTAATATACGAA 2563  
Db 244755 TGATCATCTTTTATTTATTTATCGGACCAATTTTGAATTTATCT----- 244711  
Qy 2564 AAAAATAATGATAATGCTGTTTGAATTTTATAAGAAATGATCGAGCTTTTGAATAC 2623  
Db 244710 ---TTGTTCAATAGGTTACTTTTATATACTTAAATTTTCTAGCATTAAGGAC 244655  
Qy 2624 TATTTACAAAGATTTGTTTATTTGATAGATAAATAAACAATGGTCTAAGATAAGATTG 2683  
Db 244654 ATCTCTCAAAAAAGCTTTCTCTATAATAATAGGATCGCGTATTTTGGGAGTTCTATTA 244595  
Qy 2684 GTTCTAATTTGGTTCCTCCACATGATTTTGTGGCAATTTCTTTATCAAAATGAAAAAG 2743  
Db 244594 AATCAAAATTTTGTGAATTAGATTTAATAGAAATTAAGTATGTCATTTTATAGGAT 244535  
Qy 2744 AAACAGCTTATTTTAAAGTAATCTAAATGCTCCAGATGAATATTTTATACAGACAATTA 2803  
Db 244534 GGACAAATTTATCTCGAGAAGTGA-CTTAGGTTTGGTCACTCTAACTTTATTCATAATTT 244476  
Qy 2804 TAGAAAAATAGAAATTTTCAANTAGATATCTAAATATGAAATGAA---ATTTAGATATATAA 2860  
Db 244475 TTTTGCTCTACTATTTTCTTGATATTTGTACTCAATTTATAACGACTTAAACCTGTTGT 244416  
Qy 2861 AGTCGAAAAAATCAACATCTTCTCTATTTGCTTTTACAGATGATTTCTATTCATGAATTC 2920  
Db 244415 GATGGTTTATTTTAAATTAATTTATTTGACCAATATTTATTTGACCAATATCTTTTCAAGCAAG 244356  
Qy 2921 TAAATGCAAGAAATTTTAGTTTTATTTGCTAGAAAGTTAAAAATAGAAATTAATCTA 2980  
Db 244355 GTATTATATCTGAATTTTATTTTATTTGTTACTCATTTATGTCGCAAGAAATAGCTTAATAA 244296  
Qy 2981 AA----TTTAAAGAAATTTTACTTAAAAAAATAAATAGTTGATTTTGTGAGAGTAATGTA 3036  
Db 244295 AAGAGTATTTATGAAATTTAGCACCTATGTAACAATTTTTTTTTTATTTAGTATTTTACCTTTT 244236  
Qy 3037 TGTTTAAATTTTAAATAATGACCCGGAATATTTTATTTTAAAGTACTCTCTGGTTGATTA 3096  
Db 244235 GAGTTCTACAAATTTTTTAAATTTTGTTCAAAAATTTAGATGTTCTTTTAAACAGG 244176  
Qy 3097 TTTTATTTCCAGAGCAAAAGTATGTTATTTTATTAATTTTATGAAATTTTATTTTATTTTC 3156  
Db 244175 TAGATTACATCTATCTCATTTTACAACTTTGTAGATGGTTTAACTCTTTTGGAAATAGTTT 244116  
Qy 3157 ATATPAAATTTTGAATACTAAGCTAAATATTTTAAAAAAATGAAATTTATTTTATTTAT 3216  
Db 244115 TAAGAAACAAAGTGTCTTATTTGATAATAGCTACTCTATGTTATTTAGTATGATGATGAT 244056  
Qy 3217 GGTCTATATTTATTTGTTTTCAGTAGTCACAGTATGTTTGTGGAATAAATTTTGA 3276  
Db 244055 AGTACTTACCATTGTTTGTATGTAATCTATTTATCTATAGTAAAAAGN-----TAA 244003  
Qy 3277 GATTATTTGACAGATTTTACTGCTCCCAATAATTTTGATTTTGAATATGTAATATTAAT 3336  
Db 244002 TCATAATTTGAACCTTCAACTACTCTCTATTTTAAATGTCATTAATATTTTACTGAAAGTT 243943  
Qy 3337 TGTATTTCAATTTATAAATTTGATTTATAAAAAATTAATAAATAGTATCTTTTATAGTTTTT 3396  
Db 243942 TTTATCCAGTGTGTAATGAATATAGTTGCTAGTTTGTGCTAGTAAATAATTTTGTGATG 243883  
Qy 3397 TAGTTTTATAGGTATATCTCGATTGTATATTTTCAAAATTTGGAAGAAATTTGATTTT 3456  
Db 243882 GTATCGAACCTATAAAAAAGGAATTTTACTATTTGTAATAAT--ATATGACATATTTGCTC 243825

Qy	3457	TAGACAGACACCTTATAGGACTAGACTATCTTATAACAGCGCTCAAAACCAAGGTGGTTG	3516	Db	242768	GTTATTAGAGAGTCGAGTTGAAATATTTTAAAGATTTAGAGAGCCCTAACGTTGCTCTAA	242709
Db	243824	TGATATGCGCAGGAGTAAAGAGAAATGATACCTAAAGTTATACATTTATGTTGGTTT	243765	Qy	4596	TAAATTTTCATGCTTTAAACGAGAGAGACTGTAAAAGAAATTTTGTGAGGATCTAATAT	4655
Qy	3517	GCTTTATGAACCTATCCCTACGTTAAATACCACTACAATTAGTGTTCAAATCCGTTAATCT	3576	Db	242708	TAGATATGAGGCT-----TGTTTTCTCGATATTCATTTAAACCAAGGTAATAATTG	243655
Db	243764	GGAGGAAATCCCTTACCAGATAATTTAAAGAAATATATAAAACCTTGAGAGAACAAATGT	243705	Qy	4656	AGAAAATAATGTTGGTGCAGCTTTATTCACAGATATTTATAAAGATATAAAATTTCCA	4715
Qy	3577	TTGCACCTATATAAAATAAATGCAACAATTTTTTTTCTGTGTGTCCTGTTTTATACCGA	3636	Db	242654	GTTGAATCCCAAATATTAAGAGTTAATATTCCTGATTTAAAAATATAAATATCACTTAAA	242595
Db	243704	CCGATTTATGAAATTTATGAATGAATGAGCATAAATATGATGT-----	243661	Qy	4716	AATTAATAATAGAGATTTGGTCAGGATTTGCTTTTTTAAATTTGAGGCTCTTGAACAATGT	4775
Qy	3637	TCATTTTAAAGTGGATCGAGAAATCGTAGTTTATCGCTAGCAATTAATTAATATATGCTGT	3696	Db	242594	AAATGAACTGTGTTGTTATTAATAA--TGCTTATTTAGTAGATTTTCGAATTTTCTTGT	242537
Db	243660	--TAGTAAATGTTTTATGAGAGAAGCATATCTAAGAGAAATTTTGCTTATGTTTCT	243603	Qy	4776	AACACGTGTAGTTAGTTGATACCTAGAGAAATTAATTAATAATATGCTCAATTCGTAAACAGTTC	4835
Qy	3697	TATGAGATATATAGG--TGGAATAATTTGCTTGGATATAAAAGCTAAAPAGTAATATTTGTA	3755	Db	242536	TTCGAAGGGGAGAAATTTCTTATCAGAAGAAATAATGTATATATATTTATCAAAAAGGGA	242477
Db	243602	GACTATGCAAGATTTGATATTTATATCTATCGGGGTTCTATCTAGATCTGATGTG	243543	Qy	4836	GCTTATTAATCAGAAATTCCTATATAAATAATATATGATTTAGTCACAAAGATTTGGAGAATTA	4895
Qy	3756	ATACTACTTATTTTAAATATCTGAATGCTTTTACCAATGAAATTTTGGCTGTTTATAAT	3815	Db	242476	TACTTTTGTCCCAAAATAGGAAATCTATTG---TTTTGACTACTTTAGAAGATGTTTT	242421
Db	243542	GAGCTTTTAAAGATTTAGATCCCTTTCAGGATTCATGAGTGTCTTCTAGCAAGGGAGATT	243483	Qy	4896	CCCTTTTAAAGTTAAAGAGAGTTTAGTCATTAATTTTGATGCAAAAGTTTATTAAGAGAA	4955
Qy	3816	TCTAGAGATCAAGTAAACGAGCTAGATTTATTTATCAAGGAAGTATTTGATAAGTA	3875	Db	242420	GACTTCTAATATTTTAAATATATGAGAGAAATAATTTCAAAAGGCACTAATTTTTCATTAGA	242361
Db	243482	AGTTGTGATGGAATCAGGATTAATATTTGGCGCTGTTAAAGGACATCACTTTTAAAA	243423	Qy	4956	GGTTAAATGTTTAAACAAATGTATTCAACAGATGTTTGGATAATAGAGTTCCTTGCCAAAT	5015
Qy	3876	TTAGAAAACAATATTTTATTTGGATATGGAATATCCGAATATTCAGTTACGGGAACCTGG	3935	Db	242360	AAATTCATATATGATATGTAATGTCCTTCCATAATATTTGATCGCTGGTGGTCAGC--T	242303
Db	243422	TCAAATATGCTATATATGACAAAAGTGATTAACCTCTCTTAATAAGACATGTGTAGAG	243363	Qy	5016	ATTAGAGTCTTATCGAAAAGAAATACGTAGATATCCATTTATTTAAAGCGAAAGAGATATT	5075
Qy	3936	CTCGGAAGTCATTGAGGCTATATATCATTTTTTTTATAAATCAGGAATATGTTGGGTTGATT	3995	Db	242302	ATAAGATATAAATGAACATACAGGCTTTGAAAATGACTACTTTAGGGGTAAATACCTACA	242243
Db	243362	GTTACAACTAATTTATTTGATAAACAAGGGCTTAAAGAT--AAGAAATATTATTCAAAAGA	243305	Qy	5076	ATCAAGAAAGCAATTTAGTTAGCTTGATTTTGATGAAATTTTCCGCTAAACATATATGTAAT	5135
Qy	3996	TTACTGATGTTTCTTTTATTTTATGTTATAAAAAAGTTATGAGTTAATGGGAACA	4055	Db	242242	CTAAG-----AGTGAAGAGCTGGGTGGAAATTTTGTGGGTTTTAGCTATCTT	242192
Db	243304	TTGATGATATAACAATATATCCGAGAAATTTATTTAATCAAAAGAAATTTATTAACAGGTA	243245	Qy	5136	GTTATATAAGAAATTTCAAAAGCAGTAGAGTTAAAAATGATAAAATTTAGTGTATTGTT	5195
Qy	4056	GCATATTTTATTTTACATCATTAGCCATATTTTTCATATATGAAACAATAGATCCGATT	4115	Db	242191	TGGAAGGTAGCAAGTTTAAAAAGCGAGTAAAGAAATGAATTTAATTTCAATAATAGTG	242132
Db	243244	AGGTTGATGCTGACTAGTGTACCTATTTCTATACATCATTAGAGAGAGTTGGAATA	243185	Qy	5196	CAAGTTTATAATGTAGATAAATAATTTAAAGTATGTTGTATAGAAAGCATTTAATCAAAAT	5255
Qy	4116	ATTATATATPAGPACTATTTCTTTCTTCAATAGGTATTTGGAATAATAAATTTTAAA	4175	Db	242131	CCTGTATATAAATGGAGAAATATATTTGGAGATGCTAGATAGTATCTTCTGAGCAAACT	242072
Db	243184	GTTCTTCATTTATTTTCAGATTCTCTAAAGATTAGAGTAAGGCTCATAAATTTGATTTTAT	243125	Qy	5256	TATAAAATATAGAAATATTTTGTAGATAGTGGCTCTGTAGATGATTTCTGCTAAATA	5315
Qy	4176	AAGGATATGGAGACAAAAATGAATGATTTAAATTTCAAGTTATTTGTACCAATTTATATGT	4235	Db	242071	TATCAGAAATTTAGAGATTTATCATATTTGATGATGGCTCTAGTATCGCACAGGGGATATT	242012
Db	243124	TTGATATGGTACTTTATAGAAATGCTTCTAAGGTTTCTTAAAGTTTAAAGAAATAGTTATTAT	243065	Qy	5316	TGCAAGAAATATCGCAAAAAGATAAAAAGTAGTAAAAATTTTTTTTCTACTATCATGTGGA	5375
Qy	4236	CCAAGATTATCTTGATAAATGTTATTAACAGTATTTATTAACCAACATATATAATTTAGA	4295	Db	242011	TGCGAAAAGTATTTTTTGGAGACAGCGCAATAAATAATTTCTATCAAGAAAACAGAGGC	241952
Db	243064	TGTAATAGAAGGACGAAATGTTATATATTTTACTCTTACATTTTAACAGGGCTACCGA	243005	Qy	5376	GTAATCAAAATGCTAGAAATCATGGAATAAAGCGGAGTACAGCTGAATATATTTGTTGTT	5435
Qy	4296	GGTATTTCTCGTAATGATGGAAGTACTGTAGATTTCTGAGAAATTTGCTTTAACTATAT	4355	Db	241951	CAATCAGTTGTCAGCTAACCAAGGAGTACTTAGGTGTACAGGAGATTGGAATTTGCTTTT	241892
Db	243004	TTATCCTATC--TGATGATTTCTATGTAATCAAGCAAAATAAAATTTTATATGTTGAT	242946	Qy	5436	GACTCTGATGATGCTGCTGTAGTACTAGATTAGTAGAABAAAATATATTTTAATATATATAA	5495
Qy	4356	GAAGAACGATGGAAGAATTAATATTACAGAAATTAATTAAGGAAATTTAATGAGGATGCTCG	4415	Db	241891	GACGTGATGATGATTTTATCTCTTATTTCAATTTGAGGTTATGTACAACTACAGAAAGCT	241832
Db	242945	TGTTGATGATGGTTCTCGAGGACAGTACTAAGGAAATTTGTAAGTAAATTTATATAAAGAA	242886	Qy	5496	AGTAGAAGTGAATTTATCTGGTTGTTGTACGCTACTTTTTTCAGAAAATATAAATAAATTTT	5555
Qy	4416	AAATTTTCGGACTAGAAACATGCAACAGGTAAATATATTTGCTTTGCTGATTTCTGATGACTA	4475	Db	241831	ACCAATGCCGATATTTGTTTAACTAGTATTTGGTAACTTCAACAATACTTCAATACTAGT	241772
Db	242885	TAAAGTTAGTATTTGTATATCTGTA--TAAACGTAATGGCGGTAAGCAATTCAGCCATATA	242829	Qy	5556	GAAAGTGAATTAATCCAAATATTTGATTTTGAAGCAATTAATAACCGTGCAGGACATGGAGAA	5615
Qy	4476	TATAGAAGTTGCAATTTGCGAGAGAATGCAATGATATAATCACTGAGTATAATCCCGATAT	4535	Db	241771	ATTAATTCACAGTATTTTAAAGAGATTAACCTTTATACCTTTAGAGGTGGGTTTGGAGAA	241712
Db	242828	TCTAGCAATGAGATATATGCAACCAAGTGATTAATCATGTATGCTAGATAGTATGATGTTG	242769	Qy	5616	AAAAATTTTATGAATTTGTTATATAAATAATATTTTCTCTACTCTCTGTTTGTAAACTATAT	5675
Qy	4536	AGCAGATAGATTTTGTGTTTAGTAGCAAAACGGGTATACAAAGAAAAAAGAAATAG	4595				



Ds 241711 ATGTTATATGGAAAACTTATGGAGTTTCCTCGGTTAGCAA-----AATTATATCGAGAA 241657  
Qy 5676 AAGAAAAGATACATAACAGATCTTTTTCAGAGAGATCAATGTTAGGAGAAATTTACTT 5735  
Ds 241656 GTAAATTTATTCAGTAATCCATATCCAGAAAGAAATTCATGAAGATATGATCAACATT 241597  
Qy 5736 TTTAATCTGCATTTATTAAGAAATATAGATAGAGTTAGTTATTTTGCATGAACTCTTTAT 5795  
Ds 241596 TTAAGCTTATATCTTCGCGCTTCAAAAATAGCAGTTTGTGATATTTGTCACTCGAGTGGTGT 241537  
Qy 5796 TTTTATAGCAGGTATATAAGTACAGTAATTTCTTTTAAAGAGGTGTCTTTTGGCAA 5855  
Ds 241536 ATTTTACGTAATATAGTACTACTAGGCAAAATTCATGAGAGATGT- TGTATTTTTT 241478  
Qy 5856 TTGGAATTTTGCAGAAATCAAGTATAGTATTTTAAAGCAAAATATATGTTGAGGATTTT 5915  
Ds 241477 GAGGCAATTCAGAAATATATTTGATTTTATAATCTAAATTTTCCACATAATACATCTTTA 241418  
Qy 5916 GACGTATCAATTTTAAAGATCTATAGTTGGCAAGTATTTTATATAGCTTACTAATG 5975  
Ds 241417 ATCTCTGCTGTAATATATAATCAAGTATTTGCTGGAATAGATATATGTCGGAAGATGATT 241358  
Qy 5976 TTTAAATACGAAACAGCTCTATTTTGGACAAATTTTAAATTTTAAAGATCTTTTATAA 6035  
Ds 241357 GATTTTAAATATATAGTACAGTTGATTTATACAGAAATAATATAGAAATATTTTAAAG 241298  
Qy 6036 AAATATATTTTAACTTCTTAAAGATCTAACCAAAATCTTGTCTTAAAAATTTTGT 6095  
Ds 241297 ACTATCTCTTATATAGATTTAGTGTGAAGGAGGTAAATATATATTTATTTAT 241238  
Qy 6096 ATAAGATTTGTTTCGAAACAAAGTTTAAAGAAATATTTATGTTTATATAGGAAGATATC 6155  
Ds 241237 TCCTCTATAAGATATTTCACTATTTGTAAGAAATA-----TATAATTTGAGGTATC 241185  
Qy 6156 ATGATCTATTTAGTAAATTTCTATATTTGATCTATATATATATATATATATATTTTA 6215  
Ds 241184 AAGGA-----ATGGATAAGTTAGTATATATCCAGTTTATTAACGTTCAATCTTCTTA 241129  
Qy 6216 TCTAAATGTATAGATAGCATTTGTAATCAGACCTTACAAACATATAGAGATTTCTCTGTG 6275  
Ds 241128 AACGATGTATAGAAAGTTTGTAGTCTAACTTATTTCAATTTAGAAATTTATTTCTAGTT 241069  
Qy 6276 AATCAGGTAGTACGGATAATTCGGAAGAAATTTGTTTATAGATATCGGAAGAAAGTATG 6335  
Ds 241068 AACGATGCTCTACAGATAATTCGCGGATATCTGTGATTTATTCAGAAATAGATGGG 241009  
Qy 6336 CGCATTCGTTATTTAAAGAGAAACGCGGGCTATCAGATGCGGTAATATATGCGATA 6395  
Ds 241008 CGTATATTCGTTTTCACAAAATTAATGTTGTTTATCAGATGCTCGTAATATATGGAAT 240949  
Qy 6396 AGTCGCGCAAGGTGACTACTAGCTTTTATAGACTCAGATGATTTTATTTCAATTCGAG 6455  
Ds 240948 AGTAGACTCTGGGATTTATTTATTTATCTGATTCGGAATGATTTATCTATATAAGAA 240889  
Qy 6456 TT-----ATCCCAACGTTTACAGAAACATTCAGAGAGAAATGCGCTTTGTCGAGTTGCT 6512  
Ds 240888 GATCAATCGAGAGATGTTGATTTAGCGAGAAATATATTCAGAAATGTTTATAGG 240829  
Qy 6513 GGTATG-----ATAGGTATAGTCTTCGCGGCAATTTCTTAAACAGCAGACCGCTCTCT 6566  
Ds 240828 TGTATGTTGAAAAGAGAAACATATATAAATATATCGTTTAAAGATGAAATGATA 240769  
Qy 6567 ACAATCAGGCTGTTCTGAGCGGAGGATGTTTGTAAAGAGCTGTAGAGCGGATGCT 6626  
Ds 240768 GAACTATTTAGTCAGATCAAGCAATACAAATATTTATAA-----TTATGATGCTCTATAGA 240712  
Qy 6627 CATCGCTTTGTTGGCTGTAAATAAATCTTATAAAGAAATCTTTTGAAGATTTTTCGA 6686  
Ds 240711 GCTATTTTACTGTAGCATAATAATATATATAGAGAGATTTATTTTCTACGCTGTC 240652  
Qy 6687 TTTGAAAGGTTAGATTCATGAAGATGAATCTTCACTTATCGCTTCTATGAGTTA 6746  
Ds 240651 TATCCAGTAGGGAAGTTACACGAAGATGAAATTTTAACTTATAATTTTGAAGGCC 240592

Qy 6747 GAAAAAGTTGCAATAGTAAAGAGTCTTGTACTTATTTGACCGAGAAAAATAGTATC 6806  
Ds 240591 AAAAAATATATTTCTTTTAGATAATACTTATGCATATCGTATTAGAGAAAAATAGCATA 240532  
Qy 6807 ACACTTCTAGCATGACTGACCATCGCTTCCTTCCCTACTGGAATTTCA 6856  
Ds 240531 ATGACTGGTCTCTACAATATTTAAAGGCTACATCTGTAGAAGCACTTTAA 240482  
RESULT 11  
AX954530/c  
LOCUS AX954530 349980 bp DNA linear PAT 08-JAN-2004  
DEFINITION Sequence 1376 from Patent WO03093306.  
ACCESSION AX954530  
VERSION AX954530.1 GI:40783903  
KEYWORDS  
SOURCE Streptococcus pyogenes  
ORGANISM Streptococcus pyogenes  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.  
REFERENCE 1  
AUTHORS Telford, J., Massignani, V., Margarit y Ros, I., Grandi, G., Fraser, C.  
and Tettelin, H.  
TITLE Nucleic acids and proteins from streptococcus groups a b  
JOURNAL Patent: WO 03093306-A 1376 13-NOV-2003;  
Chiron SRL (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)  
FEATURES  
Location/Qualifiers  
source 1. 349980  
/organism="Streptococcus pyogenes"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:1314"  
/note="seq 1373 to long, original length:  
2.160.266-replaced by-seq 1373: from 0.000.001 to  
0.349.980-seq 1374: from 0.300.001 to 0.649.980-seq 1375:  
from 0.600.001 to 0.949.980-seq 1376: from 0.900.001 to  
1.249.980-seq 1377: from 1.200.001 to 1.549.980-seq 1378:  
from 1.500.001 to 1.849.980-seq 1379: from 1.800.001 to  
2.149.980-seq 1380: from 2.100.001 to 2.160.266"

## ORIGIN

Query Match 11.4%; Score 798.8; DB 6; Length 349980;  
Best Local Similarity 48.2%; Pred. No. 9.7e-82;  
Matches 3323; Conservative 0; Mismatches 3382; Indels 185; Gaps 31;  
Qy 16 TGGCAATTTTGTATATCATAGCAGTTGCAATTTCTGCAATCTTAAACAGTCAATACCAA 75  
Ds 276022 TAGCGATGATTCAAACAGTTGTGTTTATTTTCTGCAAGTTTGACATTAACATTAATA 275963  
Qy 76 ATGCTGATTTAAATCG--TTCTGGAATTTTATCATAATGATGTTTCATTTATTTTGCAT 132  
Ds 275962 CTCCCAATTTTAAAGCAATAAGATTTATTTGTTCTTCTATGATCATATATATTTGCT 275903  
Qy 133 TTTTATATCTCGTATGCCAGTTGCAATTTGAGTATAGAGTAAATCTGATAGATTTGAAA 192  
Ds 275902 TTTATCTTCTGATTTTACAGAGACTTTTGGAGTCGTGGCTATCTCTGAAGAGTTTAAA 275843  
Qy 193 AAACATTTAATATAGTATATATTTTGCATTTTCTTACGGCAGTATCATTTTGTGG 252  
Ds 275842 TGGTATTTGAAATACAGCTTTTACTATATTTTTCATATCAAGTTTCATTTTATTTTATA 275783  
Qy 253 AGAATAATTTTCGCACTTTCAAGAGCTGTGCGGTATTTTACATTTAATAAATCTCGTTT 312  
Ds 275782 AAAATCTTTTACACAGCAGCTTTTCTTTTACTTTTATCTTTTATGCTATGAAATTCGATTT 275723  
Qy 313 TGGTATACCTATTATAACGTAATTTAAGCAGTTTAAAGGATAGCTTTTCTATTTTTCGACAA 372  
Ds 275722 TATTATATCTATTGAATTTCAATTTTAAATATATATCGAAATATTTCTTACCGTAAGTTT 275663  
Qy 373 TCTATCAAAAAAGACGATTTCTAATTACACGGCTGAACGATGGAATAATTCGAAGTTT 432  
Ds 275662 CACGAGATACCAAGTTGTTTTCGATAACGAATAGGATTTCTTTATCAAAAATGACCTTTA 275603

```
QY 433 TATTTGAATCACATAAACAATTCAAAAAATCTTGTTGCAATCGTAGTTTATAGGTACAG 492
Db 275602 GGAATAAAT---ACGACCAATATATATCGCTGCTGTATCTTGGATTCCTCTGAAAGG 275546
QY 493 AAATPAGATAAAATTAATTTATACCTACCTCTATTATCTGTGGAAGAAGACTATAGAGT 552
Db 275545 ATTGTTATGATTGGAACATAAATCGTTAAGGATAATAAACAAGATGCTCTTACTTCAG 275486
QY 553 TTTCAACAGGGAAGTGTGACACAGCTCTTTATAATCTACCAAGTGAGTTTTAGACG 612
Db 275485 AGTTAAGCTGCTTAACCTGTTGATCAAGCTTTTATTAACATACCACCAATTAATTTGGTA 275426
QY 613 TAAGCAATTC-----GTTTCAGATTTGAGTTGTTAGGTATTTGATGTAAGCGTTG 663
Db 275425 AATACCAATACAAGATATTATTAATGACATTTGAAGCAATGGGAGTGATTTGTCATGTTA 275366
QY 664 ATATTAAATTCATTCGGTTTTACTCGGTTTGA AAAACAAAAAATCCAACTCGTAGGTGACC 723
Db 275365 ATGTAGAGGCACCTTAGCTTTTGATAATATAGGAGAAAGCGAATCCAACTTTTGAAGGAT 275306
QY 724 ATAGCATTTGAACCTTTTCCAAAATTTTATAAGCCCTAGTCATATCATGATGAACGAC 783
Db 275305 ATAGTGTATTACATATTTCTATGAATTTCTATAAATATAGTCACCTTTATAGCAAAACGAT 275246
QY 784 TTTTGGATATACTCGGAGCGTAGTCGGTTAAATTTTGTGGTATAGTTTCTATTTTGT 843
Db 275245 TTTTGGATATACCGGGTGCTATTATAGTTTGTCTATGTGGCAATTTTTC 275186
QY 844 TAGTTCCAAATTTATTCGTAGAGATCGTGACCGGCTATTTTGTCTCAGAAAACGAGTTGGAC 903
Db 275185 TAGTTCCACAAATCAGAAAAGATGTTGACCGGCTATCTTTCTCAAAATAGATAGTGC 275126
QY 904 AGAATGAGCGATATTATACATCTTACAAAGTTTCGATCGATGTATGTTGATGCTGAGGAGC 963
Db 275125 GTAAATGGTAGGATTTTATAGATTTCTATAAATTCAGATCAATGCGAGTAGATGCAACAAA 275066
QY 964 GCRAAAGACTGCTCAGCCAAACAGACGACGAGGTGGGTATGTTTAAATGGGAA 1023
Db 275065 TTAAGAAGATTTATTTAGTTCCAAATCAAAATGACAGGG-----CTAATGTTTAAATGATGA 275011
QY 1024 AAACGATCTCAGAAATTTACTCCAAATGGAACATTTTCATACGCAAAAACAAAGTTTACAGAGT 1083
Db 275010 CGATGATCTCAGAAATTTACTAAAATAGGAAATTTATTCG-AAAAACAAGCATAGATGAGT 274952
QY 1084 TACACAGTTTATAATGTTTTTAATTTGCGGATATGAGTCTAGTTGGTACAGCTCCACTA 1143
Db 274951 TGCCTCAATTTCTAATATGTTTTTAAAGGCTGATATGAGTTTAGTAGGAACACGCGCTCCCA 274892
QY 1144 CAGTTGATGAATTTGAAAAATATATCTCTGCTCAAAAGACGAGTTGAGTTTAAACCCAG 1203
Db 274891 CAGTTGATGAATATGAAAAGTATTAATTTCAACGCAAGAGCGAGCTTAGTTTANGCCAG 274832
QY 1204 GGATTACAGGCTCTGCGCAGGTTAGTGTGCTAGTAAATATACACAGACTTCGACGACGTAG 1263
Db 274831 GAATCACTGTTTGTGGCAATATCTGTTGAGAAATATATTTACTGATTTTGAATGAAATCG 274772
QY 1264 TTCGGTTGGACTTAGCATATGATTAATTTGGAATCTCTGCTGATAGATTTAAATTTTAT 1323
Db 274771 TAAAGTTAGATGTTTCAATATATCAATGAATGCTTATTTGGTTCAGATATTAAGATTTATTC 274712
QY 1324 TAAGACAGTCAAAAGTTGTTATGTTGAGAGAGGAAATGATTAAGTAAAGTAT----- 1373
Db 274711 TCCTAACACTAAAGGTAGTTTTTACTCTCGGGAACAGAGCTTAAGTAAAGGTAAAGTTTGAAG 274652
QY 1374 -----ATGAAAGTTTGTGTTGTTGCTCGTTCTTACAGGGGACATTTGACTCACTTGTATTT 1426
Db 274651 GAAATATATGAATTTGTTCTGTTGTTTCAAGTTGGTGTCTATCTAGCACACTTGAACCT 274592
QY 1427 GTTAAACCGTTTTTGAAGGAAGAAAGACGTTTTTTGGGTAAACATTTGATAAAGAGGATGC 1486
Db 274591 TTTGAAACCCATTTTGGGAAAAAGAAAGATAGGTTTTTGGGTAACTTTTGATAAAGAGATGC 274532
QY 1487 AAGAAGTCTTTTGAAGATGA AAAAATGATATCCATGTTACTTTTCCAAACAAATCGCAATCT 1546
```



QY 4776 AACACGTTAGTAGTCTGATCTACTAGAGAAATATTATTATAATTAATGTCATTCGTACAGTTC 4835  
Db 271323 TTGCAAGGGGAGAAATTTCTTATCAGAAAGAAATAATGATATATATTTATCAAAAAGGGA 271264  
QY 4836 GCTTATTAAATCAGAAAATCTCTATAAATAATATTTAGTTAGTCACAAGATTTGGAGAATTA 4895  
Db 271263 TACTTTTGTCCCAAAATAGGAAATCTATTG-----TTTGGACTACTTAGAAGATGGTTT 271208  
QY 4896 CCCCTTTAAAGTTAAAAAGAGAGTTTAGTCATTTATTTGGATGCAAAAGTTATTAAGAGAA 4955  
Db 271207 GACTTCTAAATATTTTAAATATATGAGAAAAAAATTTCAAAAGGACTATATTTTTCATTAGA 271148  
QY 4956 GGTAAATGTTTAAACAAAATGATTTCAACAGATTTGTTGGATAATCAGTTCTTGCCAAT 5015  
Db 271147 AAATTCATATGATGTAATGTTCTTTCCATAATATATTTGATCGCTGGTGGTCAGC--T 271090  
QY 5016 ATTAGAGTCTTTATCGAAAAGAAATPACGTAGATATCCATTTTATTAAGCGGAAAGATATTT 5075  
Db 271089 ATAAAGATAAAATGAACATACAGGCTTTGAAATGACTACTTTAGGGGTAAATACCTACA 271030  
QY 5076 ATCAAGAAAGCATTTAGTTAGTTGATTTGATGAAATTTTCGCTAAACTATATGTAAT 5135  
Db 271029 CTAAG-----AGTGAAGAGGCTGGGTGGAATAATTTTGTGGGTTTTTAGCTATCTT 270979  
QY 5136 GTTATATAAGAAATTTCAAAAGCAGTAGAGTAAATGATAAATTTAGTGTATTTGTT 5195  
Db 270978 TGGAAAGTAGCAGGTTTAAAGAGCGAGTAAGAATGGAATTAATTTCAATAATAGTG 270919  
QY 5196 CCAGTTTATAGTATAGATAAATATTTAAAGTAGTTGATAGAAAGCATTTATTAATCAAAAT 5255  
Db 270918 CCTGTATATAGTGAATAATATTTGGGAGATGCTAGATAGTATTTCTTGACCAACT 270859  
QY 5256 TATAAAAATATAGAAATATTTATGATAGATAGGCTCTGTAGATGATTCGTAAATA 5315  
Db 270858 TATCAGAAATTTAGAGATATATCATAAATTTGATGATGCTCTAGTGTATCGCACAGGGGATAT 270799  
QY 5316 TGCAGGATATGCAAGAAAGATAAAGAGTAAATTTTTCACATCATAGTGA 5375  
Db 270798 TGCAGAAAGTATTTTGGAGACAGCGGAATAAATATTTCTATCAAGAAAAACAGAGGC 270739  
QY 5376 GTATCAAAATCTAGAAATATGGAATAAAGCGAGTACAGCTGAATATATTTATGTTTGT 5435  
Db 270738 CAATCAGTTGCACGTAAACAGGAGTACTTAGGTGTACAGAGATTTGGATGCTTTTAA 270679  
QY 5436 GACTCTGATGATGTTGTGATAGTATAGTATAGAAAAATATATTTTAAATATATATAA 5495  
Db 270678 GACAGTGATGATGTTTATCTTCTTATTTCAATGAGGTATGTACAAACATACAGAAAGCT 270619  
QY 5496 AGTAGAAGTATTTATCTGGTTGTTGTACGCTACTTTTTCAGNAATATATATATTTT 5555  
Db 270618 ACCAATGCCGATATGTTTAACTAGTATTTGGTAATCTCAACATATCTTCAATATCTAGT 270559  
QY 5556 GAAGTGAATATCCAAATATTTGATTTTGAAGCAATTAATACCGTGAGGACATGGGAGAA 5615  
Db 270558 ATTAATTTACAGATATTTAAAGAGATTAACCTTTATACTTTAGAGGTGGCGTTGGAAGAA 270499  
QY 5616 AAAAATTTATGAATTTGATATAAATAATTTTCTACTCTCTGTTGTGTAACATATAT 5675  
Db 270498 ATGTATTTATGGAAAACTTATGGAGTTTCCCGTTAGCAA-----AATTATATCCGAGAA 270444  
QY 5676 AAGAAAGATACATAACAGATCTTTTTCAGAGAAATCAATGGTTAGGAGAAAGATTTACTT 5735  
Db 270443 GTAAATTTATGAGTAATCCATATCCAGAGGAAAAATTCATAGAGATATGGATACAACTT 270384  
QY 5736 TTTAATCTGCAATTTTAAAGAAATATAGATAGAGTTAGTTATTTTGACTGAACATCTTTAT 5795  
Db 270383 TTAGCTTTATATCTTGGCTTCAAAAATAGCAGTTTGTGATATTTGTCACTGCAGTGGTGT 270324  
QY 5796 TTTTATAGGAGGTATACATAAGTACAGTAATTTCTTTTAAAGAGGTGTGTTTTCGAA 5855  
Db 270323 ATTTTAGCGATAATAGTACTACTAGGACAAAAATTCATAGAGAAATGT-TGTATTTTTTTT 270265  
QY 5856 TTGAAAAATTTGCAAAAAACAAGTATAGTATTTGTTTAAAGCAAAATATATGTTGAGGATTTT 5915

Db 270264 GAGGCAATTCAAAATTAATATTTGATTTATTAATCTAAAATTTTCCACAAATACATCTTTA 270205  
QY 5916 GACGTATCAATTTGTTAAAGATACCTATACGTTGGCAAGTATTTTATATAGCTTTACTAATG 5975  
Db 270204 ATCTCTGCTGTAATATATAATGAAGTATTTGGTGAATAGATATATGTCGGAAGATGATT 270145  
QY 5976 TTTAAATACGGAACACAGCTATTTTGTGACAAATTTTAAATTTTATAGAAATCTTTATATA 6035  
Db 270144 GATTTTAAATATATATGATACAGTTGATTTATACAGAAAAAATATAGAAAAATATTTTAAG 270085  
QY 6036 AAATATATTTTAACTTTGTTAAAGATCTTAAACAAAAATTTCTTTGTCTAAAAATTTTGT 6095  
Db 270084 ACTATTTCTTTTAAATATAGGATTTAGTGTGAGGAGAGGTTAAATATATATTTATTTAT 270025  
QY 6096 ATAAGAAATTTGTTTCGAACAAAGTTTAAAAAATATATATGTTTAAATAGGAAGATATC 6155  
Db 270024 TCCTCTATAAGATATTTCACTATTTGTAAGAAAAATA-----TATAATTTGAGGTTATC 269972  
QY 6156 ATGGATACCTATTTAGTAAAAATTTCTATAATTTGTACCTATATATATATGTAAGAAAAATATTTA 6215  
Db 269971 AAGGA-----ATGGATAAAGTTAGTATCATTTATCCAGTTTATAACGTTCAATCTTCTTA 269916  
QY 6216 TCTAAATATGATATAGATAGCATTTGTAATCAGACCTACAAACATATAGAGATTTCTTCTGCTG 6275  
Db 269915 AACGAGTATAGAAAGTGTTTAGCTCAAACTTTTCAATTTAGAAATTTTCTAGTT 269856  
QY 6276 AATGACGGTAGTAGCGGATAATTCGGAAGAAATTTGTTTAGCATATGCGAAGAAGATAGT 6335  
Db 269855 AACGATGGCTCTACAGATAATTCGCGGGATATCTGTGATTTATTTTTCAGAAATAGATGGG 269796  
QY 6336 CGCATTTGTTATTTTAAAAAGAGAACGGCGGCTATCAGATCCCGTATATATGCGCAT 6395  
Db 269795 CGTATATTCGTTTTCACAAAAATAATGTTGTTTATCAGATGCTCGTAATATTTGGAAT 269736  
QY 6396 AGTCGCGCAAGGTCACACTTACTTTTATAGACTCAGATCATTTTATTCATTCGAG 6455  
Db 269735 AGTAGAGCTACTCGCGGATTTATTTTACTTTGATTCGATGATTTATCTATATAAGAA 269676  
QY 6456 TT---CATCCAGCTTTTACACGAAGCAATTTGAGAGAGAAATGCCCTTTGTGGCAGTTGCT 6512  
Db 269675 GATGCAATCGAGAGATGTTGAAATTTAGCGAATAATAATTCAGAAATTTGTTTAGGG 269616  
QY 6513 GGTATG-----ATAGGTAGATGCTTCGGGGCATTTCTTAACAGCAGAGCCGCTTCCT 6566  
Db 269615 TGTATGTTTGAATAAGAGAACACATATTTATAATATCGTTTATAGAAATGAAATGATA 269556  
QY 6567 ACAAATCAGGCTCTCTGAGCGCAGGAATGTTTGTAAAAAGCTGCTAGAGCGGATGTT 6626  
Db 269555 GAAACTATTAGTCCAGTACAGCAATACAAATATTTTATAA---TTATGATGCTATAGA 269499  
QY 6627 CATCGCTTTTGTGGTGGCTGTATAAACTCTATAAAAAAGAACTATTTTGAAGATTTTCGA 6686  
Db 269498 GCTATTTTACTGTAGCACATAATAATATATAAGAGAAATTTATTTCTACGCTGTGC 269439  
QY 6687 TTTGAAAAAGGTAAAGATTGATGAAGATGAATCTTCACATTATCGTCTGCTATAGATTA 6746  
Db 269438 TATCCAGTAGGGAAGTTTACAGGAAGATGAAATTTTAACTTATAAATTTATATTTGAAAGCC 269379  
QY 6747 GAAAAAGTTGCAATAGTTAAAGAGTCTTGTACTATTATGTTCCACCGAGAAAAATAGTATC 6806  
Db 269378 AAAAATATATTTTCTTTAGATATAATACTTTATGATATCGTATTTAGGAAAAATAGCAT 269319  
QY 6807 ACAACTTTAGCATGACTGACCATCGCTTCCATTTGCCACTCTGGAATTTCA 6856  
Db 269318 ATGACTGGTTCTTACAATATTAAGAGCTACATGCTGTAGAACACTTAA 269269

RESULT 12  
AF355776  
LOCUS  
DEFINITION

AP355776 17596 bp DNA linear BCT 29-APR-2001  
Streptococcus agalactiae CNCTC 1/82 type IV capsular polysaccharide  
synthesis operon, complete sequence.

ACCESSION AF355776  
VERSION AF355776.1 GI:13876769  
KEYWORDS  
SOURCE Streptococcus agalactiae  
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.  
REFERENCE 1 (bases 1 to 17596)  
AUTHORS McKinnon,K., Chaffin,D.O. and Rubens,C.E.  
TITLE Direct Submission  
JOURNAL Submitted (03-MAR-2001) Department of Infectious Disease, Immunology and Rheumatology, Children's Hospital and Regional Medical Center, 4800 Sand Point Way NE, Seattle, WA 98105, USA  
FEATURES  
source  
Location/Qualifiers  
1..17596  
/organism="Streptococcus agalactiae"  
/mol\_type="genomic DNA"  
/strain="CNC 1/82"  
/db\_xref="taxon:1311"  
817..17187  
/gene="type IV capsular polysaccharide synthesis operon"  
817..822  
/gene="type IV capsular polysaccharide synthesis operon"  
834..839  
/gene="type IV capsular polysaccharide synthesis operon"  
928..2396  
/gene="cpsA"  
928..933  
/gene="cpsA"  
939..2396  
/gene="cpsA"  
/note="similar to Ltyr"  
/codon\_start=1  
/transl\_table=1  
/product="CpsA"  
/protein\_id="AAK43602.1"  
/db\_xref="GI:13876770"  
/translation="MSNHSRRQKKHSHTPLRVINFLVIFILLVVVSLFMYRHHF  
LAFHLNVIYGVWIVLIIILASFLCIKNAKIFTTIILVLASIFVATTLVGFKSTIDL  
TNLNKATSYSEIEMSVVVPKDSKITNLEAVSKLAAPVKNDTSNITDLIHKSEKGI  
SITQKTDSDYADNRYNRKNGDSQAMLVNLSIELSTPDFKSIQKIITYKIKKKI  
NRKNTHKGVDYIYIGDITFGSISTVSRSDVNIIMVTNTNTHKVLTTTPDAYVK  
IPDGGQYDKLHAGIYGVTSMTKLENLDYINDYARNFNSFLKLDLGGVTV  
YNDQAGTGHGNDPVPQGVTLNSEQALGFVRERYSLQGGDNRGRNOKVIAIINK  
LASSQVTKLNSITLSQTSVQTNMTIDINILNQLSTQRTFVESQALTGHSIG  
ELPSYMPGQALYMSIDQSLSNKSKIKNTMEE"  
2388..3133  
/gene="cpsB"  
2388..2393  
/gene="cpsB"  
2402..3133  
/gene="cpsB"  
/codon\_start=1  
/transl\_table=1  
/product="tyrosine phosphatase CpsB"  
/protein\_id="AAK43603.1"  
/db\_xref="GI:13876771"  
/translation="MIDHSIHVFDVDDPKTLSESLIBESYRQGVRIIVTSRHR  
KMPETPDDIIFKNPSIVHGEAKRFEHLQILYIGELYITSDMLEKRLKQIPILNNT  
KFLATEFSMOTSWKDIHTALSNVLMGLITPVVAHERYNALENQKREIINMGCVT  
QINSHILKQKLFNDKHKRPKGRARYFLEENLVHFVADSMENLDVRPPFLAEAYKIIC  
RDFGERANQLFIENAQSILKNHYI"  
3132..3137  
/gene="cpsC"  
3142..3834  
/gene="cpsC"  
/codon\_start=1  
/transl\_table=1  
/product="putative capsular polysaccharide chain length regulator Cpsc"  
/protein\_id="AAK43604.1"  
/db\_xref="GI:13876772"  
/translation="MNKIANTVEINIFNLKMLKKKFLITFVAIAFATAGLFYSLF  
IVTPQYSSSTRIYVNPNTNNSITAOLOAGSFLANDYKEIITSTVDLEKVISSEKL  
NYPSSOLLOKITISILKDTVISISVEDANPKMSQKLANSVREAAVSKIKAVTOVEDI  
TTLEKGLPKAPSPNKKVNLIGFIVGAGLSTIVLMGILDRVNTTEEDIEKVLGI  
TSIGIVFDLNLK"  
3833..4543  
/gene="cpsD"  
3833..3839  
/gene="cpsD"  
3845..4543  
/gene="cpsD"  
/codon\_start=1  
/transl\_table=1  
/product="capsular polysaccharide chain length regulator/exporter CpsD"  
/protein\_id="AAK43605.1"  
/db\_xref="GI:13876773"  
/translation="MTRLEIVDSKLRQAKTEEFYNAIRTNIQSGKENKILATSVR  
EGEGKSTISLALSQAQFKTLIIDADTRNSVMSGTFKATGKTLGNTLNSGNADL  
GDIICETNPRLMVVPSGKVPNPNTALLQNAFYNMIEAIKNIFDIYIITDTPPIGLVV  
DAATIANACDGFILVTQAGRIKRYVEKAKEQSGSGKFLGIILNKVNESVATYGDY  
GDYGNVGRDRKRK"  
4535..5944  
/gene="cpsE"  
4535..4540  
/gene="cpsE"  
4595..5944  
/gene="cpsE"  
/codon\_start=1  
/transl\_table=1  
/product="glucose-1-phosphate transferase CpsE"  
/protein\_id="AAK43606.1"  
/db\_xref="GI:13876774"  
/translation="MIQTVVYFSAITLITPNFKSNKOLLFVLIIHIVFVLSDF  
YRDFWGRYLEEFKMLVKYFPIFISLSFFIFKNSFTTTRLSFFFTITNLSILLYL  
LNSFLKYRYSAKFRDRTKVLITNKDSLRSKFNKTHDNTNIAVCILDSERDCL  
DLKHSNRIKDALTSCLTVDQAFINIPIELFGYQIQDIINDIEAMGVIVNVN  
VEALSFNIGEKRIQTPEGYSVITYSMKFYKSHLIAKRLDITGALIGLICGVAI  
FLVPIQIRKGGPAIFSONRVGRNRIIRFYKFSRMRVDAEOIKDILLVHNTQMTGLMPK  
LEDDPRITKGPIRKTSIDELPQFYNVLKDSLSLVGTPTPEYKYNSTQKRRLS  
FKPGITLWQISGRNNITDFEIVKLDVQYINWSIWSDIKIIILLTLKVLVLTGAK"  
5958..6417  
/gene="cpsF"  
5958..5963  
/gene="cpsF"  
5968..6417  
/gene="cpsF"  
/codon\_start=1  
/transl\_table=1  
/product="beta-1,4-galactosyltransferase enhancer CpsF"  
/protein\_id="AAK43607.1"  
/db\_xref="GI:13876775"  
/translation="MKICLVSSGGHIAHLNLLKPILEKEDRFWTFDKEDARSILRE  
SIVVHCFPTNRNVLKNTILAPKVLKERPDVIVSSGAAVPVPFYIGKLPCKT  
VYIEVDFRIDKPTLTGKLVPTDKFIVQWEMKKVYPKAINLGIP"  
6402..6908  
/gene="cpsIVG"  
6402..6407  
/gene="cpsIVG"  
6417..6908  
/gene="cpsIVG"  
/codon\_start=1  
/transl\_table=1  
/product="beta-1,4-galactosyltransferase CpsIVG"  
/protein\_id="AAK43608.1"  
/db\_xref="GI:13876776"  
/translation="MIFVTVTGHEQQFNRLIKEVRLKGTDAIDQEVFIOTGYSDPEP  
ONCWSKFSLYDDMNSYKAEIIVITHGGPATMNAVSKGKTLIVPQEQFGHVN  
HQVDFVNRKVTMTNFDIVDIERLQNVVYEGTMRPFLETNRNFIIEFKVILKELCD  
ENQ"  
6882..7992  
/gene="cpsIVH"

```
RBS      6882..6887
/ gene="cpsIVH"
CDS      6895..7992
/ gene="cpsIVH"
/ codon_start=1
/ transI_table=11
/ product="putative polyaccharide repeating unit
polymerase CpsIVH"
/ protein_id="AAK43609.1"
/ db_xref="GI:13876777"
/ translation="MKINKNSLFYIAIFLVNFFKSLGLGEGNSYKIVMPVAIFLPGI
KFLDLSYFERRKLVIFLLFIAFILNLFVHKVTRILLIIFLALKDLSLKAFSII
IGSRLLGVLLQFVKLDLIEIKYINPYRDQGFILRSDLGFGHPNFHNPFAVPELY
VTLFQKRLITIAFIDLNLYFLYQYISRTGYIIVLLFIIIVTKNNLIRKIFMIV
APIQLFLAFLCSTFIFFNSFVQKLSLLTGRNLNLAHLQDLVGLFNGSFKRFS
VLFDQSYMLLSMYGVVLTWFCMIIYIYISKVNVVLELOILLFIMSLVLPFESFYSI
VMNISWMVFGKIFCGGVDDLLQREFTWTKNN"
7875..7880
/ gene="cpsIVH"
7949..8800
/ gene="cpsIVH"
7949..8800
/ gene="cpsIVH"
/ codon_start=1
/ transI_table=11
/ product="putative glycosyltransferase CpsIUM"
/ protein_id="AAK43610.1"
/ db_xref="GI:13876778"
/ translation="MIVNESLGRQIKISIVPVNSKOVLIACVDSIRKQTVKNLEI
LVNDSGDSKELCEIRKRSDEIRIKFHTNGSGSSARNLGIYSTGDLIGFVSDSD
TIDPKMYETILLNIYEDQVDVQCNHKKIYNSGVNLYNGPEYVNLNKQDFLYEPLS
TKIFSVCSGLLSRLDALKIKFREEXEDTQYFDLIRKNANKFVLIISOPFFYNYIR
KNSTTSYSYQWDDIIDICTECYYAKDFNGFEVAFSLFGLFAYSILVANKIYVYNDY
RKTEELR"
8931..9680
/ gene="cpsIVH"
8931..8936
/ gene="cpsIVH"
8943..9680
/ gene="cpsIVH"
/ codon_start=1
/ transI_table=11
/ product="putative glycosyltransferase CpsIUM"
/ protein_id="AAK43611.1"
/ db_xref="GI:13876779"
/ translation="MIEKTMIPKKIHYCNFGNPKSLLKCIKSWEKYCPDYELIEW
NEONYDVNKITYTRQAYKEGSAFVSDYARLDIIYQYGYLDVDVELIKPIDDLNC
ESYFACELPGEVNTGLFGSISRNPFIKENMKIVEDTEYYSLKKTCAVITDILLTBOG
LKYINKIQVINGILLIYPPSPYPCPNMLTNKINIKKEETLTIHHYGRNWGGEKKIIDLGR
IKLGVHLLDKIFIGYGYRNILKAVKK"
9743..10701
/ gene="cpsIVH"
Query Match      10.6%; Score 743.4; DB 1; Length 17596;
Best Local Similarity 47.4%; Pred. No. 4.5e-75;
Matches 2926; Conservative 0; Mismatches 3176; Indels 77; Gaps 20;
QY      16 TGGCAATATTGATAGACAGTTCGAATTTCTGCAATCTTTAAACAAGTCATATACCAA 75
Db      4590 TAGCGATGATCAAAACCGTGTGGTTTATTTCTGCAAGTTTGCACATTAACATTAATTA 4649
QY      76 ATGCTGATTTAAATCG---TTCTGGAATTTTATCAATGATGGTTTCATTTTTCAT 132
Db      4650 CTCCCAACTTTAAAGCAATAAAGATTTATTTGTTGTTCTATTGATACATTTATTTGCT 4709
QY      133 TTTTATATCTCGTATGCCAGTTGCAATTTTGAGTATAGAGTAATCTCATAGAGTTTCAAA 192
Db      4710 TTATCTTTCTGATTTTACAGACACTTTTGGAGTCGTGGCTATCTTGAAGAGTTTAAAA 4769
QY      193 AAACATTTAACTATAGTATATATTTTCAATTTTCTACGGCAGTATCAATTTTGTGG 252
Db      4770 TGGTATTGAAATACAGCTTTTACTATATTTTTCATATCAAGTTCATTTTATTTTATTTTA 4829
QY      253 AGAATAATTTTCGCACTTTCAAGACGTGTGGCGGTGATTTTTCACATTAATAAATTCGTT 312
```

```
Db      4830 AAAATCTCTTTTCAACGACACGACTTTTCCTTTTACTTTTATCTATCTATGAAATTCGATTT 4889
QY      313 TGGTATACCTTATTAAAGTAATATTAAAGCAGTTTAAAGGATAGCTTTCTATTTTTCGACAA 372
Db      4890 TATTATATCTATTGAATTCATTTTAAATAATTATCGAATAATTCTTACGCTAAGTTTTT 4949
QY      373 TCTATCAAAAAAAGAGCAGTTCTAATTACACGGCTGAACGATGGGAAAAATATGCAAGTTT 432
Db      4950 CACGAGATACCAAAGTTGTTTTCATACGAATAAGGATTCCTTTATCAAAAAATGACCTTTA 5009
QY      433 TATTTGAAACACATAAACAATAATCAAAAAATCTTGTTCGATTTGGTAGTTTGGTACAG 492
Db      5010 GGAATAAAT---ACGACCAATAATTATATCGCTCTGTATCTTTGGATTCCTCTGAAAAGG 5066
QY      493 AAATAGATAAAAAATAATTTTATCAATACCGCTCTATTATTCTGTGGGAAGAGCTATAGAGT 552
Db      5067 ATTGTTATGTTTGAACAATAACTCGTTAAGGATAATAACAAGAAGATGCTCTTACTTCAG 5126
QY      553 TTTCAACAAGGGAAGTGGTCGACACGCTCTTTTATAAATCTACCAAGTAGTTTTTAGACG 612
Db      5127 AGTTAACTCTGCTTAACTGTTGATCAAGCTTTTATTAACATACCAATTTGAATTTATTTGGTA 5186
QY      613 TAAAGCAATTC-----GTTTCAGATTTTTCAGTTGTTAGGTATTTGATGTAAGCGTTG 663
Db      5187 AATACCAAAATACAAGATAATTATTAAATGACATTCGAAGCAATGGGAGTGAATTTGCAATGTTA 5246
QY      664 ATATTAAATCAATTCGGGTTTACCTGCGTTGAAAAACAAAAAATCCCAACTCTAGGTGACC 723
Db      5247 ATCTAGGACACTTAGCTTTTGATAATATAGGAGAAAAGCGAATCCAAACTTTTGAAGGAT 5306
QY      724 ATAGCAATTTGAATCTTTTTCACAAATTTTATAAGCCTAGTCATATCATGATGAAACGAC 783
Db      5307 ATAGTGTATTACATATTTCTGAAATTTCTATAAATATAGTCACCTTATATGCAAAAAACGAT 5366
QY      784 TTTTGGATATACCTCGGAGCGGTAGTCGGGTAAATTTATTTTGGGTATAGTTTCTATTTTGT 843
Db      5367 TTTTGGATATACCGGGTGTCTATTATAGGTTTGTCTCATATGATGGCATTTGTGGCAATTTTC 5426
QY      844 TAGTTCCAATTTATTCGTAGAGATGGTGGACCGGCTATTTTGTCTCAGAAAAAGAGTTGGAC 903
Db      5427 TAGTTCCACAATCAGAAAAGATGGTGGACCGGCTATCTTTTCTCAAAATAGATAGTGTCT 5486
QY      904 AGAATGGACGCAATTTTACATTTCTCAAGTTTCGATCGATGATGTTGATGCTGAGGAGC 963
Db      5487 GTAATGGTAGGATTTTGTAGATTCTATAAATTTAGATCAATTCGCGAGTAGATGCGAACA 5546
QY      964 GCAAAAAAGACTTGTCTCAGCCAAAACAGATGCAAGGTTGGGTATGTTTAAATGGGAA 1023
Db      5547 TTAAGAAAAGATTTTATTAGTTTCACAATCAATGACAGGG-----CTAATGTTTAAAGTTAGA 5601
QY      1024 AAAAGATCTAGAAATTTACTTCCAATTTGACATTTTATACGCAAAAAACAAGTTTAAAGCAGT 1083
Db      5602 AGATGATCCTAGAAATTTACTTAAAATAGGAAAAATTTATTCG-AAAAACAAGCATAGATGAGT 5660
QY      1084 TACCAAGTTTTTAAATGTTTTTAAATTTGGCGAATAGAGTCTAGTTGGTACAGCTCCACTTA 1143
Db      5661 TGCCTCAATTTCTATAATGTTTTTAAAGGTGATATGATGTTTGTAGGAAACAACGCGCTCCCA 5720
QY      1144 CAGTTGATGAAATTTGAAAAATATACCTCTCGTCAAAGAGACGATTCAGTTTAAACACAG 1203
Db      5721 CAGTTGATGAAATATGAAAAAGTATAATTTCAACGCAAGAGCGACGCTTAGTTTAAAGCAG 5780
QY      1204 GGAATACAGCTCTCTGGCAGGTTTAGTGGTCTGATGTAATATCAAGACTTCGACGACCTAG 1263
Db      5781 GAATCACTGGTTTGTGGCAAAATATCTGGTAGAAATAATATTAATCTGATTTTGTGATGAATCG 5840
QY      1264 TTCGTTGCACTTATGACATCAATTTGAAATTTGGACTATCTGGTCAGATATTTAAATTTTAT 1323
Db      5841 TAAAGTTAGATGTTTCAATATATCAATGAAATGGTCTATTTTGGTCAGATATTTAAGATTATTC 5900
QY      1324 TAAAGACAGTGAAGTTGTTATTTCTGAGAGAGGGAAGTAGTAAAGCTA----- 1372
```

Db 5901 TCCTAAACACTAAGGTAGTCTTACTCTGGGACAGGTGCTAAGTAAAGGTAAAGTTGGAAG 5960  
Qy 1373 -----TATGAAGTTGTTGGTGGTCTCTCAGGGGACATTTGACTCACTGTATTT 1426  
Db 5961 GAATATAATGAATTTGCTCTGGTTGGTTCAAGTGGTGTCTATCTAGCACACTTTGAACCT 6020  
Qy 1427 GTTAAACCGTTTGGGAAGGAAGAACGTTTGGTGAACATTTGGTAAAGAGGATGC 1486  
Db 6021 TTTGAACCCCATTTTGGAAAAGGAAGTAGGTTTGGGTAAACCTTTGATAAAGAGATGC 6080  
Qy 1487 AAGAAGTCTTTGAAGAAATGAATAAATGATCCATGTTTCTCAACCAATCGCAATCT 1546  
Db 6081 TAGAGTATCTAAGAGAGAGATGTTATATCATGTTCTCTTCAACAAACGTAATGT 6140  
Qy 1547 CATTAATTTAGTGAATAACTCTTCTTAGCTTTTCAAAATTTTACGTGATGAGAAACGAGA 1606  
Db 6141 CAAAACCTTGGTAAAAAATACTATCTTAGCTTTTAAAGTCTTAGAAAAGAAAGACCGAGA 6200  
Qy 1607 TGTATTTATTCATCTGGTGGCGGTTGCTGTCCTCTTTTACATCGGAAACATTT 1666  
Db 6201 TGTATCATATCATCTGGTGGCGGTTGACAGTACCATTTTATATTTGTAAGTTATTT 6260  
Qy 1667 TGGAGCAAGACGATTTATATTTAGTATTTGATCGAGTTAATAAATCTACATTAACCTGG 1726  
Db 6261 TGGTTGTAAGACCGTTTATATAGAGGTTTTCGACAGGATAGATAAACCAACTTTGACAGG 6320  
Qy 1727 AAAACTAGTTTATCCCGTAAACAGATATTTTATTTGTTTCCAGTGGGAAGAAATGAAGAAGGT 1786  
Db 6321 AAAATTAGTGTATCTGTAAACAGATAATTTTATTTGTTTCCAGTGGGAAGAAATGAAGAAGGT 6380  
Qy 1787 ATATCCTAAATCTATTAACCTTGGGAGTATTTTATTAATGATTTTGTAAACAGTAGGAAT 1846  
Db 6381 TTATCCTAAGCAATTAATTTAGGAGGAATTTTATTAATGATTTTGTACAGTGGGACA 6440  
Qy 1847 CATGAACAAAGTTAATCGATGATTAAGAGAGATGATTTTATGAATAAATAAATGGAAGT 1906  
Db 6441 CATGAACAGCAGTTCAACCGTCTTATTAAGAAAGTTGATAGATTAAGAAAGGACAGATGCT 6500  
Qy 1907 ATAAACGAGCAATATTTTATTAACAGGATATCTGACTATTTTCCAGATAATTTGCAAG 1966  
Db 6501 ATTGATCAAGAGGTGTTCAATTAACGGGTACTCAGACTTTGAACCTCAGAAATGTCAG 6560  
Qy 1967 TATAAAAAATTTCTCAGTTTACAAAGAAATGGAACAATAATTAACAATCAGAAAGTAGTT 2026  
Db 6561 TGGTCAAAATTTCTCATATGATGATATGAATCTTATCATGAAAGAGCTGAGATGTT 6620  
Qy 2027 ATTTGCCAGGAGGCCCGCTACTTTTATGAATTCATTTATCCAAAGGAAAAACAATTA 2086  
Db 6621 ATCAACATGGCGGTCCAGCGACGTTTATGAATGCAGTTTCTAAAGGGAAAAAATCTATT 6680  
Qy 2087 TTGTTTCTAGACAAAAAGTATGTTGCAACATGTAATGATCATCAAGTAGAGTTTGTA 2146  
Db 6681 GTGGTCTCTAGACAAGAACAGTTTGGAGAGCATGGAATAATCATCAGGTGGAATTTGTT 6740  
Qy 2147 AGAAGAAATTTTACAAGATAATAATTTTATTAAGAAATATAGATGATTTGTTGAA 2206  
Db 6741 AATAAGGTAAACAAATGATAATTTTGATATCGTTGATATTTGAAGGTACAAAT 6800  
Qy 2207 AAAATTTAGTGTCTTACGAAACTAACTTTTACATCAAAATAA---TAATTTTTTTTGT 2263  
Db 6801 GTAGTCTATGAGGGAACGATGAATCGTCCGTTTTTGAAGAACTAAACAGAAAGTAAATTTTATT 6860  
Qy 2264 GAAGATTTAAACAAATAGTTGAAAAATTTAATGAGGATCAAGAAATGAATAATAAAA 2323  
Db 6861 GAAGAAATTTAAGGTAATTAAGAGGTTGTGTGATGAAATCAATAAAAACTCTTTATT 6920  
Qy 2324 AGATGCATATTTGATAATGGCTTATCAATAATTTTCTCAGATTTTCTAGTTTGGAGAGGATAC 2383  
Db 6921 TTATTTGCAATTTTATTTAGTTAAATTTTAAATCACTAGGTTTGGAGAGGGAAGTCTC 6980  
Qy 2384 AGATATTTATCTCTCTCAGGAGAAATGCACACCATTTAGTTCTCTCAGAAATACCTGTAT 2443  
Db 6981 AACTTACAAATAGTGATTTTGTGCAATCTTCTTGTGTGGAATAAATTTTATTAGA 7040

Qy 2444 AATTAATTTTAAATATCTCAGGATTTATATGTTGAAATTTTACAAAGATGACAAAAATAT 2503  
Db 7041 TAGCCTTTATTTTGAAGAAGAAAACTCGTTATCATCTCTTTTATTTATTTATTTGGACCAT 7100  
Qy 2504 AAAGAAATAGGATATATGAACGATTTAAATGTTTACAGATTTATTTCTTAATATATCAGAA 2563  
Db 7101 TTTGAATTTATCTTTGTTCTAATAAGGTACTTTTATATTAATTAACCTTTAATTTTTTCTAGC 7160  
Qy 2564 AAAACTATGATATGATGTTTGAAGATTTTATTAAGAATGATCGAGCTTTTGAATAC 2623  
Db 7161 ATTA--AAGGATATCTCTCAAAAAAGCTTCTCTATAATAATAGGATCCGATTTTGT 7218  
Qy 2624 TATTTACAAAGATGTTGTTTATTTGATAGATAAAAAACAACGCTTAAAGATTAAGATTTG 2683  
Db 7219 GGAGTTCTATTAATCAAAATTTTGTGAAATTAGATTTAATAGAAATTAATAATATATCAAT 7278  
Qy 2684 GTTCTAATTTGGTTCCTCCACATGATTTTGTGGCAATTTCTTTTATCAAAATGAACAG 2743  
Db 7279 TTTTATAGGATGGACAATTTTCTGAGAGTGAATTTAGGTTTGTGTCATCCCTAACTTT 7338  
Qy 2744 AAACAGCTTATTTTAAAGTAATCTAAATGTCAGATGAATTTTATATACAGACAAATTA 2803  
Db 7339 ATTCTAATTTTGTGAGTAACTGTTTATATATGTAACACTTTTATAGAAACTA 7398  
Qy 2804 TAGAAAAATATGAATTTTCAAAATAGATTTCTAAATATGGAATTTAAGATATATAAAGT 2863  
Db 7399 AGATTAATACTATTTGCTTTTATTTTAACTCTAAATTTCTTCTGTATCAGTATATCTAT 7458  
Qy 2864 GGAATAAATCAACATCTTCTCTATGTTCTTACAGATGATCTAT--TGATGAATTCGTA 2922  
Db 7459 TCAAGAACTGATATATATAGTACTCTTATTTATATCTTATATATATGTTTACAAAGAT 7518  
Qy 2923 AATCAAGAAATTTAGGTTTTTATTTGCTAGAAAGTTTAAAAATAGAAAAATAATCTAAA 2982  
Db 7519 AACCTGATAGGAATAATTTTATGATGTTGCTCCGTACATCAACTGTTCTTGTAGCA 7578  
Qy 2983 TTTAAGAAATTTATCTAAAAATAAATAAGTTTGTGAGAGTAATGATGTTTA 3042  
Db 7579 TTTACTTTCTTCTGCTCTACTATTTTTCACATCAAAATTTGTTTCAAAAAATAGATAGC 7638  
Qy 3043 AATTAATTAATAGCCCGGAATATTTTATTTTAAAGTACTCTCTGTTGATTTATTTTA 3102  
Db 7639 CTTTTGACAGGTAGTTTAAACTGCTCATTTACAGCTTGTAGACGGCTTAACCTTTTTT 7698  
Qy 3103 TTCCAGAGCAAAAGTATGTTATTTTATTTTATGAAATTTAATTTTATTTTCAATAAA 3162  
Db 7699 GGAATAGTTTAAAGGACGAGTGTCTTATTTGATATAGTACTCTATGTTTATGAGT 7758  
Qy 3163 AATTTTGAACACTTAAGCTAATAATAAAAAATGAAATTTTATTTTATTTATTTGCTCTA 3222  
Db 7759 ATGTTATGTTAGTACTTCCATGTTTGTATGATAATCTATATATCTATAGTAAAAAA 7818  
Qy 3223 TATTAATGTTTGTTCAGTAGTCAAGATAGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTT 3282  
Db 7819 GTCAATGTTAGTGTAGCTCCAGATAC-----TTTTGTTTATTAATGCTATAGTATTTATTTA 7873  
Qy 3283 TTGACAGATTTTACTGCTCCCATATTTGGATTTTGCATTAATGTTATTTATTTATTTGTTAT 3342  
Db 7874 CAGAGAGTTTATCCCAAGTATAGTATTAATTTAGTTGATGTTTGTGTTGTTTGTGTTGTTGTT 7933  
Qy 3343 CATTTATAAATATTTGATTTATAAAAAATTTAAAAAATAGTATCTCTTTTTTATGTT--TTTTTATG 3400  
Db 7934 TTTGTTGGGGTGTAGATGATTTTACACGAGAGTTTCACTTGGACGGCAATAAATAATAGT 7993  
Qy 3401 TTTTATAGGTATATCTGCAATTTGTTATTTTCAAAATGGGAAAGATATGTTGTTTATAGA 3460  
Db 7994 GTAAATGTACCAGTATATAATTTGGAACAAATTTTAAATAGCTTGGTTGATTTCAATTAGA 8053  
Qy 3461 CAGACCTTATAGGACTAGACTATCTTTAATCAAGGCGTCAAAACAAGGTTGTTGGCTT 3520  
Db 8054 AAACAAACATATAAGAAATTTTGGAAATTTATCTTGTATGATGATCAACAGATGTTAGT 8113







Best Local Similarity 59.2%; Pred. No. 2.6e-73;			
Matches 1390; Conservative 0; Mismatches 916; Indels 41; Gaps 7;			
Qy	16	TGGCAATTAATGATAGTACGAGTTCGCAATTCCTGCAATCTTAAACAGTCATATACCA	75
Db	173	TAGCGATGATTCAAACAGTGTGTGTTATTTCTGCAAGTTTGACATTAACATTAATTA	232
Qy	76	ATGCTGATTTAAATCG--TTCCTGGAATTTTATCATAAATGATGGTTTCATTTTTCGAT	132
Db	233	CTCCCAATTTTAAAGCAATAAGATTTATTTGTTGTTCTATTTGATACATTTATTTGCT	292
Qy	133	TTTTTATATCTCGTATGCGAGTTCGAATTTGAGTATAGAGTAATCTGATAGAGTTTCAAA	192
Db	293	TTTATCTTTCTGATTTTACAGAGCTTTTGGAGTCTGTGGCTATCTTTGAAGAGTTTAA	352
Qy	193	AAACATTTAATCTATAGTATATAATTTGCAATTTTCTTACGCGAGTATCATTTTGTGTG	252
Db	353	TGGTATTCGAATAACAGCTTTTACTATATTTTCAATCAAGTTTCATTTATTTTATTTT	412
Qy	253	AGRATAATTCGCACTTTCAAGAGTGTGCGGTATTTTCAATTAATTAATTAATTAAT	312
Db	413	AAACTCATTTTACACGACGACTTTCTCTTTTCTTTTATTTGCTATGAATTCGAATTT	472
Qy	313	TGGTATACCTATTAAACGTAATTAAGCAGTTTAAAGGATAGCTTCTATTTTTCGACAA	372
Db	473	TATTTGATCTATGATTCATTTTAAATATATTCGAAATATTTCTTACGCTAAGTTT	532
Qy	373	TCTATCAAAAAGACGATCTTAATTAACAGGCTGAAACGATGGGAAATATGCAAGTTT	432
Db	533	CACGAGATACCAAGTTGTTTGTATAACGAATAAGGATCTTTTATCAAAAATGACCTT	592
Qy	433	TATTTGATCACATAAACAAATTCAAAATAATCTGTTGCAATGCTAGTTTGTAGTACAG	492
Db	593	GGAATAAAT---ACGACATAATTAATTTGCTGCTGCACTTGGACTCTCTGAAAAGG	649
Qy	493	AAATAGATAAAATTAATTAATTAACGCTCTATTTCTGTGGAAGAGCTATAGAT	552
Db	650	ATTGTTATGATTTGAAACATAACTCGTTAAGGATAATAACAAAGATGCTCTTACTTCAG	709
Qy	553	TTTCAACAGGGAAGTGGTCACACGCTGTTTATAATTAATTAACCAAGTGGTTTTAGACG	612
Db	710	AGTTAACTGCTTAACTGTTGATCAAGCGTTTATTAACATACCCATTTGAAATTTTGGTA	769
Qy	613	TAAAGCAATTC-----GTTTCAGATTTTGAATTTGTTAGTATTTGATGTAAGCGTTG	663
Db	770	AATACCAAAATACAAAGATTAATTAATTAATGACATTTGAAGCAATGGGAGTGATTTCAATGTTA	829
Qy	664	ATATTAATTCATTCGGTTTACTGCTTTGAAAAACAAAAAATCCAACTGCTAGTGACC	723
Db	830	ATGTAGAGGCATTTAGCTTTGATTAATATAGGAAAGAGCGAATCCAACTTTTGAAGAT	889
Qy	724	ATAGATTTGAATCTTTTCCAAATTTTATAAGCCTAGTGCATATCATGATGAACGAC	783
Db	890	ATAGTGTATTACATATCTATGAAATCTATAAATATAGTACCTTTATAGCAAAACGAT	949
Qy	784	TTTTTGGATATCTCGGAGCGTACTCGGTTTAAATTTTGGGTATAGTTTCTATTTTGT	843
Db	950	TTTTTGGATATCACGGTGTCTATTAGGTTTGTCTCATATGCGCAATTTTTC	1009
Qy	844	TAGTTTCCAAATTTTCTAGAGATGGTGACCGGCTATTTTGTCTCAGAAACGAGTGGAC	903
Db	1010	TAGTTTCCAAATCAGAAAGATGGTGACCGGCTATTTTCTCAAAATAGAGTAGGTC	1069
Qy	904	AGATGAGCGATATTTTACATTCACAGTTTCAGATGATGTTGTTGATGCTGAGGAC	963
Db	1070	GTAATGGTAGGATTTTATAGTTCTATAAATTCAGATCAATCGGAGTAGTACGAAACAA	1129
Qy	964	GCAAAAGACTTCTCAGCCAAACAGATGCAAGGTTGGGTATGTTTAAATGGAA	1023
Db	1130	TTAAGAAAGATTTATTTAGTTTCACAATCAAAATGACAGGG-----CTAATGTTTAAAGTTAGA	1184
Qy	1024	AAACGATCTAGAAATTTACTCCAAATTTGACATTTTCATACGCAAAACAAAGTTTACAGGAT	1083

Db	1185	CGATGATCTCTAGAAATTTACTAAAAATAGGAAATTTATTTTCG-AAAAACAAGCATAGATGAGT	1243
Qy	1084	TACCAAGATTTTATATATGTTTTTAAATTTGGCGATATAGTCTAGTTGGTACACGTCACCTTA	1143
Db	1244	TGCTCTCAATTTCTATATATGTTTTTAAAGGATGATAGTGTGTAGTAGGAACACGCGCTCCCA	1303
Qy	1144	CAGTTGATGAATTTTGAATAATATCTCTCGTCAAAAGACAGCATTTGAGTTTAAACCAAG	1203
Db	1304	CAGTTGATGAATATGAATAAGTATTAATCAACGCAAGACGACGCTTAGTTTAAAGCAAG	1363
Qy	1204	GGATTTACAGGCTCTCTGCGAGGTTAGTGTCTGTAGTATATATACAGACTTTCGACGACGTAG	1263
Db	1364	GAATCACTGCTTTGTGCAAAATATCTGTAGAAATATATATCACTGATTTTGTGATGAATTCG	1423
Qy	1264	TTGCTTTGGACTTATAGCATATCAATTTGGACTATCTGCTGATGATTAATAAATTTTAT	1323
Db	1424	TAAAGTTTAGATGTTTCAATATATCAATGAATTTGCTTATTTGCTGATGATTAAGATTAATTC	1483
Qy	1324	TAAAGACAGTGAAGTTGTTATTTGTGAGAGAGGGAAGTAAAGTAAAGTAAAGTAAAGTAAAGT	1372
Db	1484	TCCTAACACTAAAGGATAGTCTTACTTTGGGACAGGTGCTTAAGTAAAGTAAAGTAAAGTAAAGT	1543
Qy	1373	-----TATGAAAGTTTGTGTTGCTCGGTTCTTTCAGGGGACATTTTGACTCATCTGTATTT	1426
Db	1544	GAATATAATGAAATTTGCTGCTGCTGCTCAAGTGGTGGTCACTAGCACACTTTGAACCT	1603
Qy	1427	GTTTAAACCGTTTTTGGAGGAAGAACGTTTTTTTGGGTAAACATTTTGGTAAAGAGGATGC	1486
Db	1604	TTTGAACCCATTTTGGGAAAAAGAGATAGTGTGTTGGGTAAACCTTTTGTATTAAGAGATGC	1663
Qy	1487	AAGAAGTCTTTTCAAGAAATGAAAAATGATTCATGTTACTTTTCCCAACAAATCCCAATCT	1546
Db	1664	TAGGATTAATCTTAAGAGAGAGATTTGTATATCATTTGCTCTTTCCCAACAAACCCGTAATGT	1723
Qy	1547	CATTAAATTTAGTGAATAAATCTTTCTTAGCTTTTCAAAATTTTACGTGATGAGAAACCGAGA	1606
Db	1724	CAAAACTTGGTAAATAAATACTATCTAGCTTTTAAAGTCTTTAGAAAGAAAGAACCCAGA	1783
Qy	1607	TGTTATTAATTTTCAATCTGCTGCGCGCTGCTGCTCCCTCTCTTTTACATCGGAAACCTATT	1666
Db	1784	TGTTATCATATCATCTGCTGCGCTGTAGCAGTACCAATCTCTTTTATATTTGTAAGTTATT	1843
Qy	1667	TGAGCAACAGACATTTATATTTGAAGTATTTGATTCGAGTTTAAATAAATCTCATTTAACTGG	1726
Db	1844	TGGTTGAAGACCGTTTATATAGAGGTTTTCGACAGATGATTAACCACTTTTGACAGG	1903
Qy	1727	AAAACTAGTTTATCCCGTAAACAGATATTTTATTTGTTTTCAGTGGGAAGAAATGAAAGGT	1786
Db	1904	AAAAATTAGTGTATCTCTGAACAGATAAATTTATTTTTCAGTGGGAAGAAATGAAAGGT	1963
Qy	1787	ATATCTTAATCTATTTAACTTTGGGAGTATTTTAAATGATTTTGTAAACAGTAGGAACT	1846
Db	1964	TTATCTTAAAGCAATTAATTTAGGAGGAATTTTAAATGATTTTGTTCACAGTGGGACA	2023
Qy	1847	CATGAAACAGTTTAAATCCGATTAATAAGAGATTCATTTTATTTTGAAGAAATTTGGAAGT	1906
Db	2024	CATGAAACAGGATTTCAACCTCTTATTAAGAGAGTTGATGATTTAAAGGACAGGTGCT	2083
Qy	1907	ATAACCGACGAAATTTTATTTCAACAGGATATTTCTGACTATATTTCCAGATATTTGCAAG	1966
Db	2084	ATTGATCAAGAGTTTCAATTTCAACCGGTTTACTCAGACTTTTGAACCTCAGAAATTTGCAG	2143
Qy	1967	TATAAATAATTTCTCAGTTACAAAGAAATGGAACATATATATTAACAATCAGAAGTAGTT	2026
Db	2144	TGGTCAAAATTTCTCTCATATGATGATGAATGAACTCTTTACATGAAGAGAGCTGAGATTTGT	2203
Qy	2027	ATTTGCCACGAGGCGCGCTACTTTTATGAAATTTTATTTTCCAAAGGAAAAAACAATTA	2086
Db	2204	ATCACATGCGCGTCCAGCGAGCTTTTATGAATGAGTGTCTTAAGGGGAAAAAACAATTT	2263
Qy	2087	TTGTTTCTTAGACAAAAAAGTATGGTGAACATGTAAATGATCATCAAGTAGAGTTTGTGA	2146
Db	2264	GTGTTTCTTAGACAAAAAGTTTGGAGAGCAATGTAATTAATCATCAGGTGGATTTTGTG	2323







Db 482 TATTGTATCTATTGAAATTCATTTTTTAAATAATATTATCGAAAAATATTCTTACGCTAAGTTTT 541  
Qy 373 TCTATCAAAAAGACAGATCTTAATTACACGGCTGAACGATGGGAAAAATATGCAAGTTTT 432  
Db 542 CACGAGATACCAAGTTGTTTGTATACGAATAAGGATCTTTATCAAAAATGACCTTTA 601  
Qy 433 TATTGTAATCACATAAACAATTCAAAAAATCTTGTTGATTTGGTAGTTTGTAGTACAG 492  
Db 602 GGAATAAT---ACGACCAATAATTATATTGCTGCTGCATCTTGGACTCTCTGAAAGG 658  
Qy 493 AAATPAGATAAAATTAATTTATCAATACCGCTCTATTATTCTGTGGAAGAAGCTATAGAGT 552  
Db 659 ATTGTTATGATTTGAAACATAAATCTGTTAAGGATAATAAACAAGATGCTCTTACTTCAG 718  
Qy 553 TTTCAACAAGGAAGTGGTGCACACAGTCTTTTATAATCTACCAAGTGAGTTTTTAGACG 612  
Db 719 AGTTAAGCTGCTTAATCTGATCAAGCTTTTATTAATACATACCATTGGAATTAATTTGGTA 778  
Qy 613 TAAAGCAATTC-----GTTTCAGATTTTGAAGTTGTTAGGTATTGATGTAAGCGTTG 663  
Db 779 AATACCAAAATACAGATATTATTAATGACATGAGCAATGGGAGTGATTTGTCATGTTA 838  
Qy 664 ATATTAAATTCATTCGGTTTTTACTCGGTTGAAAAAACAATAAATCCAACTGCTAGGTGACC 723  
Db 839 ATGTAGAGGCATTTAGCTTTGATAATATAGGAGAAAGCGAATCCAAACTTTTGAAGGAT 898  
Qy 724 ATAGCATTTAACTTTTCCAAAATTTTATAAGCCTAGTCAATCATGATGAAGAACGAC 783  
Db 899 ATAGTGTATTACATATCTATGAAATTTCTATAAATATAGTCACTTATATAGCAAAACGAT 958  
Qy 784 TTTTGGATATACTCGGAGCGGTAGTCCGGTTAAATTTATTTGTTGATAGTTTCTTATTTTGT 843  
Db 959 TTTTGGATATCACGGGTGCTATTATAGTTTGTCTATATGTCATATGGCAATTTGGCAATTTTC 1018  
Qy 844 TAGTTCAAATTTATTCGTAGAGATGGTGCACGGCTATTTTGTCTCAGAAAACGAGTTGGAC 903  
Db 1019 TAGTTCCACAAATCAGAAAAGATGGTGCACGGCTATCTTTCTCAAAATAGAGTAGTGC 1078  
Qy 904 AGAATGAGCGATATTACATCTTACAAGTTTCGATCGATGATGTTGATGCTGAGGAGC 963  
Db 1079 GTAAATGTTAGGATTTTATAGTTCTATAAATTCAGATCAATCGCAGTAGATGCGAACAAA 1138  
Qy 964 GCAAAAAGACTTCTCAGCCAAACAGATGCAAGGTTGGTATGTTTAAATGGAA 1023  
Db 1139 TTAAGAAGATTTATTTAGTTCCAAATCAAAATGACAGG-----CTAATGTTAAGTTAGA 1193  
Qy 1024 AAACGATCCTAGAAATTTACTCCAAATGGACATTTTCATACGCAAAAACAAAGTTTAGACGAGT 1083  
Db 1194 CGATGATCCTAGAAATTTACTAAAATAGGAAAATTTATTTCG-AAAAACAAGCATAGATGAGT 1252  
Qy 1084 TACCACAGTTTATAATGTTTAAATTTGCGGATATGAGTCTAGTTGGTACACGTCACCTA 1143  
Db 1253 TGCCTCAATTTCTAATGTTTAAAGGTGATATGAGTTAGTAGGAACACGCGCTCCCA 1312  
Qy 1144 CAGTTGATGATTTGAAAAATATACTCTGCTCAAAAGACGATTCAGTTTAAACCCAG 1203  
Db 1313 CAGTTGATGATTAAGAAAGTATAATTTCAACGCAAGACGACGCTTAGTTTTAAGCCAG 1372  
Qy 1204 GGATTACAGGTTCTCGGCAGGTAGTGTCTGATGATATATACAGACTTCGACGACGTAG 1263  
Db 1373 GAATCATTGTTTGTGGCAATATCTGGTAGAAATATATACACTGATTTTGTATGAAATCG 1432  
Qy 1264 TTCGGTTGGACTTAGCATATGATTAATTGGACTATCTGGTTCAGATATTAAAAATTTAT 1323  
Db 1433 TAAAGTTTAGATGTTCAATATATCAATGAATGCTCTATTTGGTCAGATATTAAAGATTATC 1492  
Qy 1324 TAAAGACAGTGAAGTTGATTTGAGAGAGGAAAGTAAGTAAAGTA----- 1372  
Db 1493 TCCTAAACACTAAAGGTAGTCTTACTTGGGACAGGTGCTTAAGTAAAGGTAAAGTTGAAAG 1552  
Qy 1373 -----TATGAAAGTTGTTTGGTTCGTTCTCAGGGGGACATTTGACTCACTTGATTTT 1426  
Db 1553 GAATATAATGAAATTTGTTCTGTTGTTCAAGTGGTGTCTATCTAGCACACTTTGAACCT 1612

Qy 1427 GTTAAACCGTTTTTGGAGAGGAAGAAACGTTTTTTCGGTAAACATTTTGATAAAGAGGATGC 1486  
Db 1613 TTTGAAACCCATTTGGGAAAAAAGAGATAGTTTTGGTAAACCTTTTGTAAAGAGATGC 1672  
Qy 1487 AAGAAAGTCTTTTGAAGAAATGAAAAATGTATCCATGTTACTTTCCCAACAATCGCAATCT 1546  
Db 1673 TAGGAGTATTCTAAGAGAGAGAGATTGTATATCATCTCTTTTCCCAACAACCGTAATGT 1732  
Qy 1547 CATTAATTTAGTGAAAAATACTCTTTCTTAGCTTTTCAAAATTTTACGTTAGTGATGAGAAACAGA 1606  
Db 1733 CAAAAAATCTGGTAAAAAATACTATTCTAGCTTTTAAAGGTCCTTTAGAAAAAGAAACAGAGA 1792  
Qy 1607 TGTATTATTTCATCTGTTGGCGCGTTCGTCTCCCTCTTTTACATTCGGAACACTATT 1666  
Db 1793 TGTATCATATCATCTGGTGCCCTGTAGCAGTACCACTCTTTTATATATGTAAGTTATT 1852  
Qy 1667 TGAGCAAAAGACGATTTATATTGAAGTATTGTATCGAGTTTAAATAATCTACATTAACCTGG 1726  
Db 1853 TGGTTGTAGACCGTTTATATAGAGTTTTCGACAGATAGATTAACCAACTTTTGACAGG 1912  
Qy 1727 AAAAATAGTTTATCCCGTAAACAGATATTTTTTATTTGTTTCAGTGGGAAGAAATGAAGAAGT 1786  
Db 1913 AAAAATAGTGTATCTGTAAACAGATAAAATTTATTTGTTTCAGTGGGAAGAAATGAAAAAGT 1972  
Qy 1787 ATATCTCTAAATCTATTAACTTGGGGAGTATTTTTTAAATGATTTTGTAAACAGTAGGAACT 1846  
Db 1973 TTTATCTTAAGGCAATTAATTTAGGAGGAATTTTTTAAATGATTTTGTTCACAGTGGGACA 2032  
Qy 1847 CATGAACAACAGTTTAAATCGATTGATAAAGAGATGATTATTTGAAAAAATAATGGAAGT 1906  
Db 2033 CATGAACAGCATTTCAACCGTCTTTTAAAGAAGTTGATAGATTTAAAGGACAGGTGCT 2092  
Qy 1907 ATAACCGACGAATATTTTATTTCAACAGGATATTTCTGACTATATATCCAGAATATTCGAAG 1966  
Db 2093 ATTTGATCAAGAAGTGTTCATTCAACCGGTTACTCAGACTTTTGAACTTCAGAAATTTGTCAG 2152  
Qy 1967 TATAAAAAATTTCTCAGTTTCAAAAGAAATGGAACAATATATTAAACAATCAGAAGTAGTT 2026  
Db 2153 TGGTCAAAAATTTCTCTCATATGATGATATGAACCTTTACATGAAGAAGCTGAGATTGTT 2212  
Qy 2027 ATTTGCCACGAGGCCCGCTACTTTTATGAAATTCATTATCCAAAGCAAAAAACAATTA 2086  
Db 2213 ATCACAATGCGCGTCCAGCGAGCTTTATGAAATGCGAGTTTCTAAAGGGAATAAACTATT 2272  
Qy 2087 TTGTTTCTTAGACAAAAAAGTATGTTGAACATGTAAATGATCATCAAGTAGAGTTTGTGA 2146  
Db 2273 GTGGTTCTTAGACAAGAAAGTTTGGAGAGCATGTGAATAATCATCAGGTGGATTTTGTG 2332  
Qy 2147 AGAAGAAATTTTACAAGATAATAATATTTTATTTATAGAAAATATAGATGATTTGTTGAA 2206  
Db 2333 AAAAGAGTTATTCTTGAATAATGAATAGATTATATTGTAATATCAGTGAATTAGAGAA 2392  
Qy 2207 AAAATATTGAAGTTTCTTAAGCAAC---TAACTTTACATCAAAATAATAATTTTTTTTGT 2263  
Db 2393 ATTATTAAAGAAAAAATAATATCTACTAGTAAAGTATATACAAAACAATGATTTTGT 2452  
Qy 2264 GAAAGATTTAAACCAATAGTTGAAAAATTTTAAATGAGGATCAAGAAAATGAATAATAAAA 2323  
Db 2453 TCCTCTTTCAAAAATGAACCTTCTTAAACTATTTTGAATAATAATATTTTGTGGAGAAAAA 2512  
Qy 2324 AGATGCA 2330  
Db 2513 AATTGAA 2519

RESULT 16  
AF332893  
LOCUS  
DEFINITION  
Streptococcus agalactiae strain 090 capsular polysaccharide  
biosynthesis gene cluster, partial sequence.  
AF332893  
AF332893.1 GI:18033311  
ACCESSION  
VERSION



## KEYWORDS

SOURCE Streptococcus agalactiae  
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.

REFERENCE 1 (bases 1 to 4888)

AUTHORS Kong, F., Gowan, S., Martin, D., James, G. and Gilbert, G.L.

TITLE Serotype identification of group B streptococci by PCR and

sequencing

J. Clin. Microbiol. 40 (1), 216-226 (2002)

JOURNAL

MEDLINE

PUBMED

21635022

11773119

2 (bases 1 to 4888)

AUTHORS Kong, F. and Gilbert, G.L.

TITLE Direct Submission

Submitted (02-JAN-2001) CIDMLS, ICPMR, Westmead Hospital, Darcy

Rd., Sydney, NSW 2145, Australia

JOURNAL

FEATURES

Location/Qualifiers

1..4888

/organism="Streptococcus agalactiae"

/mol\_type="genomic DNA"

/strain="090"

/serotype="Ia"

/db\_xref="taxon:1311"

<1..135

/gene="cpsaD"

<1..135

/gene="cpsaD"

/codon\_start=1

/transl\_table=11

/product="CpsaD"

/protein\_id="AAL57052.1"

/db\_xref="GI:18033312"

/translation="AKQMEQSGSKFLGIILNKVSESATVVDGYDGYNGYKDRKREK

"

361..1536

/gene="cpsaE"

361..1536

/gene="cpsaE"

/codon\_start=1

/transl\_table=11

/product="CpsaE"

/protein\_id="AAL57053.1"

/db\_xref="GI:18033313"

/translation="MWLYKSPYIFISSLFFIFKNSFTTTLRLSFFSFIAMNSILLVL

LNPLKYRYKYSAKFSRDTKVLIITKDSLSKMTFRKVDHNYVIANCILDSEKDCY

DLKNSRLINKDALTSELCTLDQAFINIPILFKYQIQDIINDIEAMGVIVYN

VEALSFONIGKRIQTEGVSIVITYSMKFYKYLIAKRFIDITGAILGLICGIVAI

FLVPQIRKGGPAIFQSNVRNGRIIFRFYKFSRMVDAEQIKDILLVHQMOTGLMFK

LDDPRITKTKGKIRKTSIDELPOFYNVLKGDMSLVGTRPTVDVEYKYNSTQKRRLS

FKPGITGLWQISGRNNITDPDEIVKLDVQYINWSIDIKIILLTLKVLVLTGAK"

1560..2009

/gene="cpsaF"

1560..2009

/gene="cpsaF"

/codon\_start=1

/transl\_table=11

/product="CpsaF"

/protein\_id="AAL57054.1"

/db\_xref="GI:18033314"

/translation="WKICLVGSGGHLAHLNLLKPIWEKEDRFVWTFDKEDARSILRE

EIVVHCFFPTNRNKNLIVKNTILAFKVLKRPDVIISGGAIVAVPFYIGKLPCKGT

VYIEVFDRIKPTLTKLVPTVDKFIQVQEMKVPKALNLGGIF"

2009..2491

/gene="cpsaG"

2009..2491

/gene="cpsaG"

/codon\_start=1

/transl\_table=11

/product="CpsaG"

/protein\_id="AAL57055.1"

/db\_xref="GI:18033315"

/translation="WIFVVGTHQEQFNRLIKEVDRLKGTGAIDQEVFIQTGYSDPEP

"

gene

CDS

2587..3726

/gene="cpsaH"

2587..3726

/gene="cpsaH"

/codon\_start=1

/transl\_table=11

/product="CpsaH"

/protein\_id="AAL57056.1"

/db\_xref="GI:18033316"

/translation="WQGLLLALIVLLICSSYNEKMKFLNMBIFFVFIYIILTSIL

LHSLFKNPDFRILAAFNLSLIGVSLAFHYKNTLELDKILKAFLENGLIPLFLG

GTYCYLHNNIQNISIFGRDLGSDWINGMHTQAMGFYFYSNLIIPITVVTNLYIYI

YMLRYSIMVTIGVLLFTFILPITGSRAGIVAILAQMFTLLNTLVVKKTKFLL

YILPFLAVIMVLYFDNLLSYRIINLRSGSSSRFSLYKDTNIVNNLLKGEV

KELWNSDLPLGSHSTYGYFYKSGLLGMNIVPEGLLLIFNIGRKAQKQSAFYIEVG

TLITLFFFALEDLDGANWLIVFTVLGILENKKDFYSQLKRWKS"

3726..4700

/gene="cpsaI"

3726..4700

/gene="cpsaI"

/codon\_start=1

/transl\_table=11

/product="CpsaI"

/protein\_id="AAL57057.1"

/db\_xref="GI:18033317"

/translation="MEKRLIVSIIPIYNSEAYLKECVSLQOOTHPLIEVILLDDGS

TDNSGEICDNLSDQEDNRILVFRKNGGVSSARNLGDSTGEFITFVDSDDFVAPNMI

EIMKLNLITNADIAEVDNISERDYRKKRRNFYKFKNNLSKFLSGNRVNI

CTLYKXSIIIGNLRFDENLIGEDLFNCLLCOEHRIVVDVTSLSYTYRIVKTSAMN

QKFNESLDPTTFPNEVSSLPVAKLANVEAKPLREKIKLKRKMFELGSNDNKKIKVQ

RELFPKDKSVPPFYKAYKYLKLLSFLYLMKCSPLKYVMAYRFRQK"

4734..>4888

/gene="cpsaJ"

4734..>4888

/gene="cpsaJ"

/codon\_start=1

/transl\_table=11

/product="CpsaJ"

/protein\_id="AAL57058.1"

/db\_xref="GI:18033318"

/translation="MIKKIBKDLISVIVPIYNVEDYLVECTIESLIVOTYRNIILLIN

DGSDTNCA"

ORIGIN

Query Match 10.4%; Score 729.4; DB 1; Length 4888;

Best Local Similarity 59.2%; Pred. No. 2.6e-73;

Matches 1390; Conservative 0; Mismatches 916; Indels 41; Gaps 7;

Qy 16 TGGCAATTATTGTATGATAGACAGTTGCAATTTCTGCAATCTTAAACAGTCATATACCA 75

Db 182 TAGCGATGATTCAACAGTTGTGGTTATTATTCTGCAAGTTTGACATTAACATTAATTA 241

Qy 76 ATGCTGATTTAAATCG--TTCTGGAATTTTATCAATATCATGATGTTTATTTTGCAT 132

Db 242 CTCCCAATTTTAAAGCAATAAGATTATTGTTGTTCTATTGATATCATTAATTATGTC 301

Qy 133 TTTTATATCTCGTATGCCAGTTGAAATTGAGTATAGAGGTAATCTGATAGAGTTTGAAA 192

Db 302 TTTATCTTCTGATTTTACAGACACTTTTGGAGTCGTGGCTACTCTTGAAGAGTTTAAA 361

Qy 193 AAACATTAATAGTATGATTAATATTGCAATTTTCTTACGGAGTATCATTTTGTGG 252

Db 362 TGGTATTGAAATACAGCTTTTACTATATTTTTCATATCAAGTTTCATTTATTTTATTA 421

Qy 253 AGAATAATTTCCGCACTTTCAAGCGTGTCCGCGTATTTTCCACATTAACTTCGTTT 312

Db 422 AAAACTCATTTTACAAACGACGACTTTCCTCTTTTCTTTTATTTGCTATGAATTCGATT 481

Qy 313 TGGTATACCTATTAAACGTAATTAATAAGCAGTTTAAAGGATAGCTTCTTCTATTTTCGACAA 372

Db 482 TATTGTATCTATTGAAATTCATTTTAAATAATATATCGAAATAATATCTTACGCTAAGTTTT 541  
Qy 373 TCTATCAAAAAGACAGATCTCTAATTAACAACGGCTGAACGATGGGAAAAATATGCAAGTTTT 432  
Db 542 CACGAGATACCAAGTGTGTTGATAACGAATAGGATCTTTATCAAAAAATGACCTTTA 601  
Qy 433 TATTGTAATCACATAPAAACAATTCAAAAAATCTTGTTGCATCTGGTAGTTTATAGGTACAG 492  
Db 602 GGAATAAAT---ACGACATAATATATATGCTGTCATCTTGGACTCTCTGAAAGG 658  
Qy 493 AAATPAGATAAAATTAATTTATCATTTACCGCTCTATTAATCTGTGGAAGAAGCTATAGAGT 552  
Db 659 ATTGTTATGTTGAACATAAATCTGTTAAGGATAAACAAGAAGTCTCTTACTTTCAG 718  
Qy 553 TTTCACAAGGAAGTGGTCGACACGCTCTTTATAAATCTACCAAGTGAAGTTTTTAGACG 612  
Db 719 AGTTAACTCGTTAACTGTTGATCAAGCTTTTATTAACATACCAATTTGAATTTATTTGGTA 778  
Qy 613 TAAAGCAATTC-----GTTTCAGATTTTGAAGTGTGTTAGGTATTTGATGTAAGCGTTG 663  
Db 779 AATACCAAAATACAAGATTAATTAATGACATTTGAAGCAATGGGAGTAGTGTCAATGTTA 838  
Qy 664 ATATTAATTCATTCGGTTTTACTCGGTTGAAAAACAAAAATCCAACTGCTAGGTGACC 723  
Db 839 ATGTAGAGGCACCTTAGCTTTGATAATATAGGAGAAAGCGAATCCAACTTTTGAAGGAT 898  
Qy 724 ATAGCATGTAACTTTTTCACAAAATTTTATAAGCCTAGTCATATCATGATGAACGAC 783  
Db 899 ATAGTGTATTACATATCTATGAAATCTTATAAATATAGTCACCTTTATAGCAAAACGAT 958  
Qy 784 TTTTGGATATACCTCGAGCGGTAGTCGGGTTAATTTTGTGTTGATATGTTTCTATTTTGT 843  
Db 959 TTTTGGATATACCGGGTGTCTATATAGGTTTGTCTATATGGCATTTGTGCGCAATTTTC 1018  
Qy 844 TAGTTTCCAAATATTTCGTAGAGATGGTGCACGGCTATTTTGTCTCAGAAAACGAGTTGGAC 903  
Db 1019 TAGTTTCCAAATACAGAAAGAGTGGTGCACGGCTATCTTTTCTCAAAATAGAGTAGGTC 1078  
Qy 904 AGAATGACGCATATTTCATTTACAAAGTTTCGATCGATGTATGTTGATGCTGAGGAGC 963  
Db 1079 GTAATGCTAGGATTTTATAGTTCTATATAATTCAGATCAATGCGAGTAGATGCGAAACAA 1138  
Qy 964 GCAAAAAGACTTCTCAGCCCAACCAACGATGCAAGGTGGGTATGTTTAAATGGGAA 1023  
Db 1139 TTAAGAAAGATTTTATAGTTTACAAATCANATGACAGGG-----CTAATGTTAAGTTAGA 1193  
Qy 1024 AAACGATCCTAGAAATTAATCTCAATTTGACATTTTCATACGCAAAAACAAAGTTTACAGCAGT 1083  
Db 1194 CGATGATCCTAGAAATTAATAGGAAATTTATTCG-AAAAACAAGCATAGATGAGT 1252  
Qy 1084 TACACAGTTTATAATGTTTAAATTTGCGGATATGAGTCTAGTTGGTACACGTTCCACCTA 1143  
Db 1253 TGCCTCAATTTCTAATAATGTTTAAAGGTGATATGAGTTAGTAGGAACACGCGCTCCCA 1312  
Qy 1144 CAGTTGATGAATTTGAAAATATACCTCTGCTCAAAAGACGATGAGTTTAAACCAG 1203  
Db 1313 CAGTTGATGAATTAAGAAAGTAAATTAATTCACGCAAGAGACGCGCTTAGTTTAAAGCCAG 1372  
Qy 1204 GGATTAACAGTCTCTGCGAGGTTAGTGGTCTGTAGTAATATACAGACCTTCGACGACGTAG 1263  
Db 1373 GAATCACTGGTTGTGGCAATATCTGTAGAAATATATATCATGATTTTGTAGAAATCG 1432  
Qy 1264 TTCGTTGGACTTAGCATATCAATTTGATAATTTGGACTATCTCGTCAGATATTTAAATTTAT 1323  
Db 1433 TAAAGTTAGATGTTCAATATATCAATGAATGGTCTATTTGGTCAGATATTAAGATTATTC 1492  
Qy 1324 TAAAGACAGTGAAGTGTATGTTGAGAGGGAAGTAAAGTA----- 1372  
Db 1493 TCCTAAACACTAAAGGTAGTCTTACTTGGGACAGGTGCTAAGTAAAGGTAGGTTTGAAG 1552  
Qy 1373 -----TATGAAAGTTTGTGTTGGTCTTCTCAGGGGGACATTTGACTCACTTGATTTT 1426  
Db 1553 GAATATAATGAAAATTTGTTCTGGTTGGTTCAAGTGGTGGTCACTCTAGCACACTTGAACCT 1612

RESULT 17  
AF363058  
LOCUS  
DEFINITION  
Streptococcus agalactiae strain Fp6453 capsular polysaccharide  
biosynthesis gene cluster, partial sequence.  
AF363058  
AF363058.1 GI:18029163  
VERSION

Qy 1427 GTTAAAAACCGTTTTTGGAGGAAGAAAGAAACGTTTTTGGGTAAACAATTTGATTAAGAGGATGC 1486  
Db 1613 TTTGAAACCCATTTTGGGAAAAAAGAGATAGGTTTTTGGTAACTTTTGTATTAAGAGATGC 1672  
Qy 1487 AGCAAGCTCTTTGGAAGAATGAAAATATGATCCATGTTACTTTTCCCAACAATCGCAATCT 1546  
Db 1673 TAGGAGTATTTCTAAGAGAAGAGATTGTTATATCATGCTTCTTCTTCCCAACAACCGTAATGT 1732  
Qy 1547 CATTAATTTAGTGAAAAATACCTTTCTTAGCTTTCAAAAATTTTACGTGATGAGAAACCGA 1606  
Db 1733 CAAAACTTGGTAAAAAATACATTTCTAGCTTTTAAAGTCCCTAGAAAAGAAAGACCAGA 1792  
Qy 1607 TGTATATTTCACTCTGGTGGCGCGTGTGTCGCCCTCTCTTTTATCATCGGAAAAATTTT 1666  
Db 1793 TGTATATCATATCTCTGGTGGCGCTGTAGCAGTACCATCTCTTTATATTTGTAAGTTATT 1852  
Qy 1667 TGGAGCAAGACGATTTATATTTGAAGTATTTGATCGAGTTTAAATAATCTACATTAACCTGG 1726  
Db 1853 TGGTTGTAAGACCGTTTTATATAGAGGTTTTTCGACAGATAGATAAAACCAACTTTTGACAGG 1912  
Qy 1727 AAAAAGTAGTTTATCCCGTAAACAGATATTTTTTATTTGTTTCAGTGGGAGAAAATGAAGAAGT 1786  
Db 1913 AAAAATTAGTGTATCTGTAACAGATAAATTTATTTGTTTCAGTGGGAGAAAATGAAGAAGT 1972  
Qy 1787 ATATCCTAAATCTATTAATCTTGGGAGTATTTTTTAAATGATTTTGTAAACAGTAGGAAT 1846  
Db 1973 TTATCCTAAGGCAATTAATTTTAGGAGGAAATTTTAAATGATTTTGTACAGTGGGACCA 2032  
Qy 1847 CATGACAAACAGTTTAAATCGATTTGAAGAGATTTGATTTTATGAAAAAATTTGGAAGT 1906  
Db 2033 CATGAAACAGAGTTTCAACCGCTTTATTAAGAAAGTTGATAGATTAAAGGAGCAGGTGCT 2092  
Qy 1907 ATAAACGAGCAATATTTTATTTCAACAGGATATTTCTGACTATATTTCCAGATAATTTGCAAG 1966  
Db 2093 ATTGATCAAGAGTGTTCATTCAACGGGTACTCAGACTTTGAACTTCAGAAATTTGTCAG 2152  
Qy 1967 TATAAAAAATTTCTCAGTTTCAAAAGAAATGGAAACAATATATTAACAATCAGAAAGTAGTT 2026  
Db 2153 TGGTCAAAAATTTCTCTCATATGATATGAACTCTTTACATGAAAGAAAGCTGAGATTGTT 2212  
Qy 2027 ATTTGCGACGAGGCGCCGCTACTTTTGAATTTTCAATTTATCCAAAGGAAAAACAATTA 2086  
Db 2213 ATCACACATGGCGGTCCAGCGAGTTTATGAATGCACTTTCTAAAGGAAAAACAATTT 2272  
Qy 2087 TTGTTTCTTAGACAAAAAAGTATGCTGAACATGTAATCATCAAGTAGAGTTTGTGTA 2146  
Db 2273 GTGGTTCTTAGACAAAGAACAGTTTGGAGAGCATGTGATATCATCAGGTGGATTTTGTG 2332  
Qy 2147 AGAAGAAATTTTACAAGATAATAATATTTTATTTATAGAAAATATAGATGATTTGTTGAA 2206  
Db 2333 AAAGAGTTATTTCTTGAATAATGAATTAGATTATTTTGAATATCAGTGAATTTAGAGAAAT 2392  
Qy 2207 AAAATTTATGAAGTTTCTAAGCAAC---TAACTTTTACATCAAAATATAATTTTTTTTGT 2263  
Db 2393 ATTATTAAAGAAAAAATATCTCTAGTAAAGTAAATATCACAACAATGATTTTGT 2452  
Qy 2264 GAAAGATTAAAAAATAATGTTGAAAAATTTTAAATGAGGATCAAGAAAATGAATATAAAAA 2323  
Db 2453 TCCTCTTTCAAAAATGAACTTTCTTAACTATTTTGAATAATAATATTTTGTGGAGAAAA 2512  
Qy 2324 AGATGCA 2330  
Db 2513 AATTGAA 2519

KEYWORDS Streptococcus agalactiae  
SOURCE Streptococcus agalactiae  
ORGANISM Bacteria: Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.  
1 (bases 1 to 4888)  
REFERENCE Kong, F., Gowan, S., Martin, D., James, G. and Gilbert, G.L.  
AUTHORS Serotype identification of group B streptococci by PCR and  
TITLE sequencing  
J. Clin. Microbiol. 40 (1), 216-226 (2002)  
JOURNAL  
MEDLINE 21635022  
PUBMED 11773119  
REFERENCE 2 (bases 1 to 4888)  
AUTHORS Kong, F. and Gilbert, G.L.  
TITLE Direct Submission  
JOURNAL Submitted (22-MAR-2001) CIDMLS, ICPMR, Westmead Hospital, Darcy  
Road, Sydney, NSW 2145, Australia  
FEATURES  
source  
1..4888  
/organism="Streptococcus agalactiae"  
/mol\_type="genomic DNA"  
/strain="FP6453"  
/serotype="1a"  
/db\_xref="taxon:1311"  
<1..135  
/gene="cpsD"  
<1..135  
/gene="cpsD"  
/codon\_start=1  
/transl\_table=11  
/product="CpsD"  
/protein\_id="AAL56381.1"  
/db\_xref="GI:18029164"  
/translation="AKEQEQSGSKFLGILNKNVSVATYGDYGDYGYKDRKRK  
"  
361..1536  
/gene="cpsE"  
361..1536  
/gene="cpsE"  
/codon\_start=1  
/transl\_table=11  
/product="CpsE"  
/protein\_id="AAL56382.1"  
/db\_xref="GI:18029165"  
/translation="WVLKYFYVIFLSSFLFKNSFTTTLSPSFAMNSILLYL  
LNSFLKRYKYSYAKRSRTKVLIINKDSLSKWTFRNKYQIDNYIAVCILDSSEKCY  
DKHNSLRILINKDALSELCTVDQAFINIPILFKGYQIDINDIISAMGVIVNVN  
VVALSPDNTIGEKRIQTFEGYSVITYSKFYKSHLAKRFLDITGAILGLICGVAI  
FLVPOIKDGGPAIFQNRVGRNGRIFRFYKFRSMRVDARQIKDLVHNMGTGLMFK  
LDDPRITKIGKFIKTSIDELPOFYNVLKGDMSLVGTRPTVDEYKYNSTQKRRLS  
FRGITGLQISGRNNITDFDELIVKLDVQYINWSIWSDIKILLTLKVLLGTGAK"  
1560..2009  
/gene="cpsF"  
1560..2009  
/gene="cpsF"  
/codon\_start=1  
/transl\_table=11  
/product="CpsF"  
/protein\_id="AAL56383.1"  
/db\_xref="GI:18029166"  
/translation="MKICLVGSGGHLAHLNLIKPIWEKEDRFWTFDKEDARSILRE  
EIVYHCFPTNRNVKNLVKNTILAPKVLARKRPDVIISGAAVAVPFFYIGKLFGCKT  
VIVLEFDRIDKPLTKLVYPTDKFIVQWEEMKVPYPRAINLGGIF"  
2009..2491  
/gene="cpsG"  
2009..2491  
/gene="cpsG"  
/codon\_start=1  
/transl\_table=11  
/product="CpsG"  
/protein\_id="AAL56384.1"  
/db\_xref="GI:18029167"  
/translation="MIFVTVGTHEQQFNRLIKEVDRLKGTGAIDQBEVFIQTGYSDPEP  
"

QNCQWSKFLSYDDMNSYMKBAEIVITHGGPATFMNAVSGKKTIVVPRQEQGEHVNN  
HQVDFLKELFKYELDYILINSELENIIEKRNISTSKVISQNNDFCSPFNLSKLFE  
"  
2587..3726  
/gene="cpsH"  
2587..3726  
/gene="cpsH"  
/codon\_start=1  
/transl\_table=11  
/product="CpsH"  
/protein\_id="AAL56385.1"  
/db\_xref="GI:18029168"  
/translation="MQLLLALLIVLLICSSNEMKFLNMAEIFFIVFIIVLTSL  
LHSLKPNDFDRIILAAFNLSIIGIVSLAFHYHYKNTTLELDKILKAFNLGLLFLFG  
GTYVCLHNNIQNISIFGRDLIGSDWINGMHTQRAMGFYFYSNLIPIVVTNLYIYI  
YMLKRNYSIMTIGVLLFTFPIIPIGSGSRAGIVAILAOMFILLANTVVKKTKIFLL  
YMLPILVIVMMLYFDNLLSIYIRINLRSGSSSESRFSLYKDTNIVINNSLLFGGV  
KELWNSDLPGLGSHSTYIGFYKSGLLGLMNVLPGLLLIFTNIGRAKQSAFYEIVG  
TLTILFSFPALEDLDGANLIVFIITVLIGLENKDFYQLKRWKS"  
3726..4700  
/gene="cpsI"  
3726..4700  
/gene="cpsI"  
/codon\_start=1  
/transl\_table=11  
/product="CpsI"  
/protein\_id="AAL56386.1"  
/db\_xref="GI:18029169"  
/translation="MEKRILVSIPIYNSEAYLKECVQSLQOOTHLEVLILDDGS  
TDSGEICDNLQEDNRILLFFKKGCVSSARNLGLDKSTGEFTFVDSDDFVAPNMI  
EIMLNLTENADIAEVDIFDSNERDYRKKRRNFYKFNKNNLSKEFLSGNKNENIV  
CTKLYKSGIIGNLRDENLKI GEDLLFNCKLCOEHRIVVDTSSTLYTRI VKTSAMN  
QKFNESDIFITIFNEVSLVPAKLANYVEAKFLREKIKLRMFELGNSINDNKIKVQ  
REIFFNDIKSYFPYKAVKYLKLLSLFYLKCSPLKYVMAYRRFQKQ"  
4734..>4888  
/gene="cpsJ"  
4734..>4888  
/gene="cpsJ"  
/codon\_start=1  
/transl\_table=11  
/product="CpsJ"  
/protein\_id="AAL56387.1"  
/db\_xref="GI:18029170"  
/translation="MIKTEKDLISVIVPYNVEDYLVBCIESLIVQTYRNIIEILLIN  
DGSITDNCA"  
ORIGIN  
Query Match 10.4%; Score 729.4; DB 1; Length 4888;  
Best Local Similarity 59.2%; Pred. No. 2.6e-73;  
Matches 1390; Conservative 0; Mismatches 916; Indels 41; Gaps 7;  
QY 16 TGGCATTATTGATATGATAGCAGTTGCAATTCCTGCAATCTTAAACAGTCATATACCA 75  
DB 182 TAGCGATGATTCAACAGTTGTGGTTTATTTTCTGCAAGTTTGACATTAACTAATTA 241  
QY 76 ATGCTCAATTAATCG---TTCGGAAATTTTATCATATGATGATGGTTCAATTTTTCAT 132  
DB 242 CTCGCCAATTTTAAAGCAATTAAGATTATTTGTTGTTCTATTGATACATTATATGTCT 301  
QY 133 TTTTATATCTCGTATGCCAGTTGAATTTGAGTATAGAGTAATCTGATAGAGTTTGAAA 192  
DB 302 TTTATCTTCTGATTTTACAGAGACTTTTGGAGTCGTGGCTATCTTGAAGAGTTTAAAA 361  
QY 193 AAACATTTAATATAGTATAATATTTGCAATTTTCTTACGGCAGTATCATTTTGTGG 252  
DB 362 TGGTATTGAATACAGCTTTTACTATATTTTTCATATCAAGTTTCATTTTATTTTATA 421  
QY 253 AGAATAATTTCCGACTTTTCAAGACGGTGGCGGTGATTTTTCATATAATAAATCTCGTTT 312  
DB 422 AAAACTCATTTTAAACGACGACGACTTTCCTTTTTTCTTTTATGATGAGTTCGATTT 481  
QY 313 TGGTATACCTATTTAACGTAATTTATTAAAGCAGTTTAAAGATAGCTTTCTATTTTCGACA 372

Db 482 TATTGTATCTATTGAATTCATTTTAAATAATTTATCGAAAAATATCTTACGCTAAGTTTT 541  
Qy 373 TCTATCAAAAAGACGATCTTAATTTACAACGGCTGAACGATCGGAAAAATATGCAAGTTT 432  
Db 542 CACGATACCAAGTGTGTTTGTATACGAATAAGGATCTTTATCAAAAAATGACCTTTA 601  
Qy 433 TATTTCGAATCACATAAACAATAAATCAAAAAATCTCTGTTCATTTGGTAGTTAGGTACAG 492  
Db 602 GGNATAAT---ACGACCAATAATATATGCTGTCTGCACTCTGGACTCTCTGAAAAGG 658  
Qy 493 AAATAGATAAAATTAATTTATCAATACCGCTCTATTAATTTCTGTGGGAAGAGCTATAGAGT 552  
Db 659 ATTGTTATGATTTGAAACATAACTCTGTAAAGATATAACAAGATGCTCTTACTTCAG 718  
Qy 553 TTTCAACAAGGAGTGTGCGACACGCTCTTTATAAATCTACAAGTAGTTTTTAGAGG 612  
Db 719 AGTTAACTGCTTAACGTGTGTGATCAAGCTTTTATTAACATACCCATTGAATTAATTTGGTA 778  
Qy 613 TAAAGCAATTC-----GTTTCAGATTTTGAGTTGTAGGTATTTGATGTAAGCGTTG 663  
Db 779 AATACCAATAACAGATATTAATGACATTAAGCAATGGAGTGATTTGCAATGTTA 838  
Qy 664 ATATTAATTCATTCGGTTTTTACTGCGTTGAAAAACAAAAAATCCAACTGCTAGGTGACC 723  
Db 839 ATGTAGAGGCACCTAGCTTTTGATAATATAGGAGAAAGCGAATCCAACTTTTGAAGGAT 898  
Qy 724 ATAGCATTTGTAATTTTCCACAATAATTTTATAAGCTTAGTCAATATCATGATGAACGAC 783  
Db 899 ATAGTGTATACATATCTATGAAATCTATAAATATAGTCACTTATAGCAAAACGAT 958  
Qy 784 TTTTGGATATCTCGAGCGGTAGTCGGGTAAATATTTTGGTATAGTTTCTATTTTGT 843  
Db 959 TTTTGGATATCAGGGTGCTATATAGGTTTGTCTATATGTCATATGTCGCAATTTGGCAATTTTC 1018  
Qy 844 TAGTTCCAAATTAATTCGTAGAGATGTGGACCGGCTATTTTGTCTCAGAAACAGATTGGAC 903  
Db 1019 TAGTTCCAAATCAGAAAGATGTGGACCGGCTATCTTTCTCAAATAGTAGGTCT 1078  
Qy 904 AGAATGGACGATATTTTCAATTTCAAGTTTCGATCGATGATGTTGATGCTGAGGAGC 963  
Db 1079 GTAATGGTAGGATTTTGTAGATTTCTATAAATTCAGATCAATGGGATAGATGCAAGCAAA 1138  
Qy 964 GCAAAAAGACTTGCTCAGCCAAACAGATGCAAGGTGGGTATGTTTAAATGGGAA 1023  
Db 1139 TTAAGAAAGATTTATTTAGTTCAAAATCAAAATGACAGG-----CTAATGTTTAAAGTTAGA 1193  
Qy 1024 AAACGATCTAGAAATPACTCCAATTGGACAATTTCAACGCAAAAACAAGTTTAGACGAGT 1083  
Db 1194 CGATGATCTAGNAATTAATAATAGGAAATTTATTTCG-AAAACAAGCATAGATGAGT 1252  
Qy 1084 TACCACAGTTTTTAATGTTTTTAATTTGGCGATATAGTCTAGTTGGTACACGTCACCCCTA 1143  
Db 1253 TGCCTCAATCTATAATGTTTTAAAAAGGTGATATGAGTTTAGTAGGAAACACGCCCTCCCA 1312  
Qy 1144 CAGTTGATCAATTTGAAAATATCTCCGTTGCAAGAGAGATTTGAGTTTAAACCG 1203  
Db 1313 CAGTTGATGAATATGAAAAGTAAATTTCAACGACAGAAAGCGCCCTTAGTTTAAAGCCAG 1372  
Qy 1204 GGATACAGGCTCTCGGACAGGTTAGTGGTCGTAGTAAATACAGACTTCGACGACGAGTAG 1263  
Db 1373 GAATCACTGGTTGTGGCAAAATCTGGTAGAATAATATCACTGATTTTGATGAATCG 1432  
Qy 1264 TTCGGTTGGACTTAGCATATGATGATTAATTTGGAATCTAGTCTAGATATTAATAATTTTAT 1323  
Db 1433 TAAAGTTAGATGTTCAATATATCAATGAATGCTCTATTGTTGGTCAGATATTAAGATTAATTC 1492  
Qy 1324 TAAAGCAGTGAAGCTGTTATTTGAGAGAGGAGTAAGTAAAGTA----- 1372  
Db 1493 TCCTAACACTAAAGGTAGTCTTACTTTGGACAGGTCGCTAAGTAAAGGTAAAGTTGAAAG 1552  
Qy 1373 -----TATGAAGTTTGTGTCGGTCTTTCAAGGGGACATTTGACCTCAGTCTGTATTT 1426  
Db 1553 GAATATAATGAAAAATTTGTCGTGTTTCAAGTGGTGTCTATAGCACACTTGAACCT 1612

Qy 1427 GTTAAAAACGTTTTTGGAGGAAGAAAGCGTTTTTTGGGTAAACATTTTGTATAAAGAGGATGC 1486  
Db 1613 TTTGAAACCCATTTTGGGAAAAAGAGATAGGTTTTTTGGGTAAACCTTTGTATAAAGAGATGC 1672  
Qy 1487 AAGAAAGTCTTTTGAAGAATGAAAAAATGTATCCATGTTTACTTTTCCAAACAAATCGCAATCT 1546  
Db 1673 TAGGAGTATTTCAAGAGAAGAGATTTGTATATCATTTGCTCTTTTCCAAACAAACCGTATGTT 1732  
Qy 1547 CATTAATTTAGTGAATAAATACCTTCTTACGTTTTCAAAATTTTTACGTGATGAGAAACCGA 1606  
Db 1733 CAAAAACCTTGTGTAATAAATACTATTTCTAGCTTTTAAAGGTCCTTAGAAAAAGAAAGACCGA 1792  
Qy 1607 TGTATATTTTCACTCTGGTGGCGCGTGTGCTGCCCTTTCTTTTATCATCGGAAAAACTATT 1666  
Db 1793 TGTATATCATATCATCTGGTGGCGCTGTAGCAGTACCAATCTTTTATATTTGTTGTAAGTTATT 1852  
Qy 1667 TGGAGCAAGACGATTTATATTTGAAGTATTTGATCGAGTTATTAATCTACATTAACCTCG 1726  
Db 1853 TGGTTGTAAGACCGTTTATATAGAGTTTTCGACAGGATAGATAAACCAACTTTTGACAGG 1912  
Qy 1727 AAAACTAGTTTTATCCCGTAACAGATATTTTATTTGTTTCAGTGGGAAGAAATGAAAGAGT 1786  
Db 1913 AAAATTTAGTGTATCTGTAAACAGATAAATTTATTTGTTTCAGTGGGAAGAAATGAAAGAGT 1972  
Qy 1787 ATATCCTAAATCTATAAATTTGCGGAGTATTTTTTAATGATTTTTTTGTAACAGTAGAACT 1846  
Db 1973 TTATCCTAAGGCAATTAATTTTAGGAGGAAATTTTTTAATGATTTTTTGTACAGTGGGACA 2032  
Qy 1847 CATGAACAACAGTATTAATCGATTTGAATAAGAGATTTGATTTATTCGAAAAAATGCAAGT 1906  
Db 2033 CATGAACAGCAGTTTCAACCGTTTATTAAGAAAGTTGATTAAGAGGACAGGTGCT 2092  
Qy 1907 ATAAACCGAATAATTTATTCAAAACAGGATATTTCTGACTATATTTCCGAAATATTGCAAG 1966  
Db 2093 ATTTGATCAAGAAGTGTTCATTCAAACGGGTACTCAGACTTTGAACTTCAGAAATTTGTCAG 2152  
Qy 1967 TATAAAAAATTTCTCAGTTTCAAAAGAAATGGAAACAATATATTAACAATCAGAAAGTAGTT 2026  
Db 2153 TGGTCAAAAATTTCTCTCATATGATGATATGAACCTTTACATGAAAGAGCTGAGATTTGT 2212  
Qy 2027 ATTTGCCAGCGGCGCCGCTACTTTTATGATTTATTTATTTATTTTCAAAAGGAAAAACAATTA 2086  
Db 2213 ATCACACATGGCGGTCCAGCGCAGTTTATGAATGAGTTTCTTAAAGGGAAAAAACTATT 2272  
Qy 2087 TTGTTTCTTAGACAAAAAAGTATGCTGAAACATGTAATAATGATCATCAAGTAGAGTTTGT 2146  
Db 2273 GTGGTTCTTAGACAAGACAGTTTGGAGAGCATGTGAATAATCATCAGGTGATTTTTTG 2332  
Qy 2147 AGAAGAAATTTTCAAGATAAATAATATTTTATTTATAGAAAAATATAGATTTGTTTGA 2206  
Db 2333 AAAGAGTTTATTTCTTGAATAATGAATTTAGATTTATATTTTGAATATCATGTAATTAGAAT 2392  
Qy 2207 AAAATTTTGAAGTTTCTAAGCAAAAC---TAACTTTTACATAAATAAATTTTGT 2263  
Db 2393 ATTATTAGGAAAAAATAATATCTACTAGTAAGTAATATCAAAACAAATGATTTTGT 2452  
Qy 2264 GAAAGATTTAAACAATAATAGTTTGAATAATTTAATGAGGATCAAGAAAAATGAATAAATAA 2323  
Db 2453 TCCCTTTCAAAATGAACCTTTCTAATACTATTGTAATAATATATTTTGTGGAGAAAA 2512  
Qy 2324 AGATGCA 2330  
Db 2513 AATTGAA 2519  
  
RESULT 18  
AF363057  
LOCUS  
DEFINITION  
Strepococcus agalactiae strain GB00/017 capsular polysaccharide  
bioynthesis gene cluster, partial sequence.  
ACCESSION  
AF363057  
VERSION  
AF363057.1 GI:18029155

KEYWORDS	Streptococcus agalactiae	gene	2587..3726	ONCOMSKPLSYDDMNSYKMEABIVITHGGPATFMAVSKGKKTIVVPQEOFGHEVNN
SOURCE	Streptococcus agalactiae	CDS	/gene="cpsH"	HQVDFLKLFLKELDYILNISELENIKERNIISTKVISQNDPFCSSFKNELSKLFE
ORGANISM	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;		2587..3726	
REFERENCE	Streptococcus		/gene="cpsH"	
AUTHORS	1 (bases 1 to 4888)		/codon_start=1	
TITLE	Kong,F., Gowan,S., Martin,D., James,G. and Gilbert,G.L.		/transl_table=11	
	Serotype identification of group B streptococci by PCR and		/product="CpsH"	
	sequencing		/protein_id="AAL56378.1"	
JOURNAL	J. Clin. Microbiol. 40 (1), 216-226 (2002)		/db_xref="GI:18029160"	
MEDLINE	21635022	gene	/translation="MQLLLLALIVLLICSSYNEKMKFLNMAEIPFIVFYIYLSIL	
PUBMED	11773119	CDS	/db_xref="GI:18029160"	LHSLFKPDPDFRIILAAFNLSLIGVSLAFYHYKNTTLELDKILKAPLFGNLPLFLG
REFERENCE	2 (bases 1 to 4888)		3726..4700	GTYYCLHNNIQTISIFGRDLIGSDWINGMHTQAMGPFYSNLIPIITVVTNLVIYI
AUTHORS	Kong,F. and Gilbert,G.L.		3726..4700	YMLKRNYSIMTIGVLLFTFLPIGSGRAGIVAILAQMIFILLNTLVVVKKTKFLV
TITLE	Direct Submission		/gene="cpsI"	YILPLAVIMVLYEDNLISLYRILNLRSGSSBSRSLYKDTNIVNLNLSFGEV
JOURNAL	Submitted (22-MAR-2001) CIDMLS, ICPMR, Westmead Hospital, Darcy		/codon_start=1	KELWNSDPLGSHSTYIGYFKYKGLGLNMNIVPGLLI1FTNIGRKAQKQAPFYIEVG
	Road, Sydney, NSW 2145, Australia		/transl_table=11	TLITLFSFFALDELGDGANLIVFTFTVLGILENDKDFYSQLKRWKS"
FEATURES	Location/Qualifiers		/product="CpsI"	
source	1..4888		/protein_id="AAL56379.1"	
	/organism="Streptococcus agalactiae"	gene	/db_xref="GI:18029161"	
	/mol_type="genomic DNA"	CDS	/translation="MEKRILVSIILPIYNSAYLKECVQSVLQOHPLELIVLIDGGS	
	/strain="GB00/017"		TDSGEICDNLSDQEDNRILVPHKNGGVSSARNLGDSTGEFTTFVDSDDFVAPNMI	
	/serotype="Ia"		EIMLNLTENADIAEVDNISERNYRKKRRNFYKFKNNLSLKEFLSGNRVENIV	
	<1..135		/product="CpsI"	CTKLYKKSIIIGNLAFDENLIGEDLFCNKLCOEHRIVVDVTTSLTYRIVKTSAMN
	<1..135		/protein_id="AAL56374.1"	QKFENSLDFTITFNEVSSILVPAKLANYVEAKFLREKIKLRKMPGLSGNIDNKIKVQ
	/gene="cpsD"		/db_xref="GI:18029156"	REIFPKDKSPFYFYKAVKYLKGLLSFYLKMCSPKLYVMAYRRFQK"
	/gene="cpsD"		/translation="MIKKIEKDLISVIPIYVNVEDYLVECIIESLIVQTYRNIETILLN	
	/codon_start=1	gene	4734..>4888	DGSTDNCA"
	/transl_table=11	CDS	/gene="cpsJ"	
	/product="CpsD"		4734..>4888	
	/protein_id="AAL56375.1"		/gene="cpsJ"	
	/db_xref="GI:18029157"		/codon_start=1	
	/translation="MWLVKSPYIPISSLPFIFKNSPTTTLRSLFSPFIAMNSILLYL		/transl_table=11	
	INSFLKYKYSYAKFSRDTKVLTITNKSLSKMTFRNKYDHNIVAVICILDSSEKCY		/product="CpsJ"	
	DKNSLRILNKDALTSCLTVDQAFINIPLELFKYIQIDINDIEANGVIVN		/protein_id="AAL56380.1"	
	VEALSPNIGBKRIQTFEGYSVITYSMKFKYSHLIAKRFIDITGAILIGLICGVAL		/db_xref="GI:18029162"	
	FLVPQIRKDGPAIFSQNRVGRNGRIFFYFRSMRVDAEQIKKDLVHNMQTMGLMF		/translation="MIKKIEKDLISVIPIYVNVEDYLVECIIESLIVQTYRNIETILLN	
	LDDPRTIKIGKFIKRTSIDELPOFYNVLKGDSILVGRPTTVDYEKYKNSGTQRRLS		4734..>4888	
	FKPGITGIWQISGRNITDFDEIVKLDVQYINWSIDSIKILLTLKVLVLTGAK"		/gene="cpsJ"	
	1560..2009	gene	/gene="cpsJ"	
	/gene="cpsP"	CDS	4734..>4888	
	1560..2009		/gene="cpsJ"	
	/gene="cpsP"		/codon_start=1	
	/transl_table=11		/transl_table=11	
	/product="CpsP"		/product="CpsJ"	
	/protein_id="AAL56376.1"		/protein_id="AAL56380.1"	
	/db_xref="GI:18029158"		/db_xref="GI:18029162"	
	/translation="MKICLVGSGGHLAHLNLLPIKEWEDRFWVTFPKEDARSLIRE		/translation="MIKKIEKDLISVIPIYVNVEDYLVECIIESLIVQTYRNIETILLN	
	BIVVHCPRPTNRNKNLTKTILAPKVLKERPDVLISSGAAVAPFPYVIGKLGCKT		4734..>4888	
	VYIEVFDRIQFTLTGKLVYPVTDKFIQWHEMKKYKPAINLGGIF"		/gene="cpsJ"	
	2009..2491	gene	/gene="cpsJ"	
	/gene="cpsG"	CDS	4734..>4888	
	2009..2491		/gene="cpsJ"	
	/gene="cpsG"		/codon_start=1	
	/transl_table=11		/transl_table=11	
	/product="CpsG"		/product="CpsJ"	
	/protein_id="AAL56377.1"		/protein_id="AAL56380.1"	
	/db_xref="GI:18029159"		/db_xref="GI:18029162"	
	/translation="MIFVTVGTHEQQFNRLIKEVDRLKGTGAILDQEVFIQTGYSDPEP		/translation="MIKKIEKDLISVIPIYVNVEDYLVECIIESLIVQTYRNIETILLN	

Db 482 TATTGTATCTATTGAATTCATTTTAAATAATATCGAATAATATCTTACGCTAAGTTTT 541  
Qy 373 TCTATCAAAAAGACAGATCTTAATTAACAAGCGCTGAACGATGGGAAAAATATGCAAGTTT 432  
Db 542 CACGAGATACCAAAAGTTGTTTGTATAACGAATAAGGATCTTTATCAAAAATGACCTTTA 601  
Qy 433 TATTGTGATCACAATAAACAATAACAAAATCTGTTGTCATGCTAGTATTAGGTACAG 492  
Db 602 GGAATAAAT---ACGACCAATAATATATGCTGCTGCAATCTTGGAATCTCTGAAAGG 658  
Qy 493 AAATAGATAAAATTAATTTATCATTTACCGCTCTAATATTCTGTGGAAGAAGCTATAGAGT 552  
Db 659 ATTGTTATGATTTGAAACATAAATCGTTAAGGATAATAAACAAGATGCTCTTACTTCAG 718  
Qy 553 TTTCAACAGGGAAGTGGTCGACACAGCTCTTTATAAATCTACCAAGTGAGTTTTTGAAGC 612  
Db 719 AGTTAACTGCTTAACCTGTTGATCAAGCTTTTATAAATACATACCAATTTGAATTTTGGTA 778  
Qy 613 TAAAGCAATTC-----GTTTCAGATTTTGAAGTTGTTAGGTATTGATGTAAGCGTTG 663  
Db 779 AATACCAATAACAAGATATTAATGACATTTGAAGCAATGGGAGTGATTTGTCATATGTTA 838  
Qy 664 ATATTAAATTCATTCGGTTTTTACTCGGTTGAAAAACAAAAAATCCAACTGCTAGGTGACC 723  
Db 839 ATGTAGAGGCACCTTAGCTTTGATATAATAGGAGAAAGCGAATCCAACTTTTGAAGAT 898  
Qy 724 ATAGCATGTAACTTTTTTCCAAATTTTATAAGCCTAGTCATATCATGATGAAACGAC 783  
Db 899 ATAGTGTATTACATATTTCTATGAATTTCTATAAATATAGTCACCTTATAGCAAAACGAT 958  
Qy 784 TTTTGGATATACTCGGAGCGTAGTCGGTTAAATTTATTTGGGTATAGTTTCTATTTTGT 843  
Db 959 TTTTGGATATACCGGTTGCTATTATAGTTTGTCTATATGTCATATGTCGCAATTTTTC 1018  
Qy 844 TAGTTCCAAATTTATCGTAGAGATGGTGACCGGCTATTTTGTCTCAGAAACGAGTTGGAC 903  
Db 1019 TAGTTCCAAATCAGAAAGATGGTGACCGGCTATCTTTTCTCAAAATAGAGTAGGTC 1078  
Qy 904 AGAATGGACGATATTTACATCTTACAAGTTTTCATGATGATGTTGATGCTGAGGAGC 963  
Db 1079 GTAATGGTAGGATTTTATAGATTTCTATAAATTCAGATCAATCGAGTAGATGCAACAACA 1138  
Qy 964 GCRAAAGACTCTCGACCAAAACAGATGCAAGGTGGGTATGTTTAAATGGAA 1023  
Db 1139 TTAAGAAAGATTTATATAGTTCACAATCAAAATGACAGGG-----CTAATGTTAAGTTAGA 1193  
Qy 1024 AAACGATCCTAGAATTTACTCCAATTTGACATTTTCATACGCAAAAACAAAGTTTACAGCAGT 1083  
Db 1194 CGATGATCCTAGAAATTAATAAATAGGAAATTTATTCG-AAAAACAAGCATAGATGAGT 1252  
Qy 1084 TACCACAGTTTATTAATGTTTAAATTTGGCGATATGAGTCTAGTTGGTACAGCTCCACTA 1143  
Db 1253 TGCCTCAATTTCTAATAGTTTAAAGGTGATATGAGTTTAGTAGGAACACGCGCTCCCA 1312  
Qy 1144 CAGTTGATGATTTGAAAAATATCTCTGCTCAAAAGAGACGATTCGAGTTTAAACCAAG 1203  
Db 1313 CAGTTGATGATATGAAGAATTAATTTCAACGACAGAGCGCGCTTAGTTTAAAGCCAG 1372  
Qy 1204 GGATTACAGGTCTCTGGCAGGTTAGTGGTCTAGTAAATATACACAGACTTCGACGACGTAG 1263  
Db 1373 GAATCAGCTGGTTTGGCAATAATCTGGTAGAATATATCATCTGATTTTGTGATGAATCG 1432  
Qy 1264 TTCGGTTGGACTTAGCATAATGATTAATTTGGACTATCTGGTCAGATTAATAAATTTTAT 1323  
Db 1433 TAAAGTTAGATGTTTCAATATATCAATGAATGGTCTATTGTCAGATTAAGATTTATTC 1492  
Qy 1324 TAAAGCAGTCGAAGTTGTTTGTGAGAGGAGGAGTAAAGTA----- 1372  
Db 1493 TCCTAAACCAATTAAGGTAGTCTTACTTTGGGAACAGGTGCTTAAGTAAGGTAAAGTTGAAAG 1552  
Qy 1373 -----TATGAAAGTTGTTTGGTGGGTTCTTTCAGGGGGAATTTGACTCACTTGTATTT 1426  
Db 1553 GAATATAATGAAATTTGCTGTTGGTTTCAAGTGGTGGTCACTCTAGCACACTTGAACCT 1612

Qy 1427 GTTAAAAACCGTTTGGAGGAAGAACGTTTTTGGGTAAACATTTGATTAAGAGGATGC 1486  
Db 1613 TTTGAAACCCATTTTGGGAAAAAGAGATAGTTTGGGTAACTTTGATTAAGAGATGC 1672  
Qy 1487 AGAAGTCTTTTGAAGAAATGATAATCTCATGTTACTTTCCAAACAATCCCAATCT 1546  
Db 1673 TAGAGTATTTCTAAGGAGAGATTTGTATATCATTTGCTTCTTTCCAAACAACCCGTAATGT 1732  
Qy 1547 CATTTAATTTAGTCAAAAATACTTTCTTAGCTTTTCAAAATTTTACGTGATGAGAAACGAGA 1606  
Db 1733 CAAAAACCTTGGTAAAAATACTATTCTTAGCTTTTAAAGTCTCTTAGAAAAAGAAACGACAGA 1792  
Qy 1607 TGTATTATTTCTATCTGCTGCGCGCTTCTGCTCCCTTCTTTTACATCGGAAACTATT 1666  
Db 1793 TGTATCATATCATCTGGTCCGCTGTAGCAGTACCATTCTTTTATATTTGTAAGTTATT 1852  
Qy 1667 TGGAGCAAGACGATTTATATTCGAAGTATTGTATTCGAGTTTAAATAATCTACATTAACCTGG 1726  
Db 1853 TGGTTGTAGACCGTTTATATAGAGGTTTTCGACAGATAGATAAACCACTTTGACAGG 1912  
Qy 1727 AAACTAGTTTATCCCGTAAACAGATATTTTTTATTTGTTTCAGTGGGAAGAAATGAAGAAGGT 1786  
Db 1913 AAAATTAGTGTATCCCTGTAACAGATAAATTTATTTGTTTCAGTGGGAAGAAATGAAAAAGT 1972  
Qy 1787 ATATCCTAAATCTATTAACTTTGGGAGTATTTTTTAATGATTTTGTAAACAGTAGGAACT 1846  
Db 1973 TTTATCCTTAAGGCAATTAATTTAGGAGGAATTTTTTAATGATTTTGTTCACAGTGGGACA 2032  
Qy 1847 CATGAAACAACAGTTTAACTCGATTTGATAAAGAGATTTGATTTATTTGAAAAAAAATGGAAGT 1906  
Db 2033 CATGAACAGCGTTTCAACCGCTCTTTTAAAGAGATTTGATAGATTAAAGGACAGGTGCT 2092  
Qy 1907 ATAACCGACGAAATATTTATTTCAACAGAGTATTTCTGACTATATTTCCAGAAATATGCAAG 1966  
Db 2093 ATTGATCAAGAAGTGTTCATTCAACCGGTTTACTCAGACTTTTGAACCTCAGAAATTTGCAG 2152  
Qy 1967 TATAAAAAATTTCTCAGTTACAAAGAAATCGAAACATATATTATAAATCAGAAGTAGTT 2026  
Db 2153 TGGTCAAAATTTCTCTCATATGATGATTAATGAACTCTTTACATGAAGAGAGCTGAGATTGTT 2212  
Qy 2027 ATTTGCCACGAGCGCCGCTACTTTTATGAAATTCATTTATCCAAAGGAAAAACAATTA 2086  
Db 2213 ATCACAATGCGCGTCAGCGAGCTTTTATGAATGCGAGTTTCTAAAGGGAAAAAATCTATT 2272  
Qy 2087 TTGTTTCTTAGACAAAAAAGTATGTTGAACATGTAAATGATCATCAAGTAGAGTTTGTGA 2146  
Db 2273 GTGGTTCTTAGACAAGAACAGTTTGGAGAGCATGTGAATAATCATCAGGTGGAATTTTGTG 2332  
Qy 2147 AGAAGAAATTTTACAGATATAATATTTTATTTATAGAAAAATATAGATGATTTGTTGAA 2206  
Db 2333 AAGGAGTTATTCTTTGAAATATGAATTAGATTATATTTTGAAATATCAGTGAATTAGAAT 2392  
Qy 2207 AAAATTATTCAGTTTCTTAAGCAAAAC---TAACTTTACATCAAAATAATAATTTTTTTTGT 2263  
Db 2393 ATTATTAAAGAAAAAATATATCTACTAGTAAAGTATATATCAAAACAATGATTTTTTGT 2452  
Qy 2264 GAAAGATTAAAAACAATAGTTGAAAAATTTAATGAGGATCAAGAAATATGAATAATAAAA 2323  
Db 2453 TCCTCTTTTCAAAAATGAACITTTCTAAACTATTTTGAATAAATAATATTTTGTGGAGAAAAA 2512  
Qy 2324 AGATGCA 2330  
Db 2513 AATTGAA 2519

RESULT 19  
AF363059  
LOCUS  
DEFINITION  
Streptococcus agalactiae strain SG99/276 capsular polysaccharide  
biosynthesis gene cluster, partial sequence.  
AF363059  
AF363059  
VERSION  
AF363059.1 GI:18029171

KEYWORDS	Streptococcus agalactiae	gene	2587. 3726	ONCWSKFLSYDDMNSYKMEABIVITHGGPATFMNAVSKGKTTIIVPRQEQGEHVN
SOURCE	Streptococcus agalactiae	CDS	/gene="cpsH"	HQVDFLKEFLKELYDILNATISELNIKERNIISTKVISQNDNFCSSFKNELSKLFE
ORGANISM	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;		2587. 3726	"
REFERENCE	Streptococcus		/gene="cpsH"	
AUTHORS	1 (bases 1 to 4888)		/codon_start=1	
TITLE	Kong, F., Cowan, S., Martin, D., James, G. and Gilbert, G.L.		/transl_table=11	
	Serotype identification of group B streptococci by PCR and		/product="CpsH"	
	sequencing		/protein_id="AAL56392.1"	
JOURNAL	J. Clin. Microbiol. 40 (1), 216-226 (2002)		/db_xref="GI:18029176"	
MEDLINE	21635022		/translation="MQLLLLALVLLICSSYNKMKFLNMAEIFFVIFYIYLTSL	
PUBMED	11773119		/db_xref="GI:18029176"	LHSLFKNPFDRILAFNLSLIGIVAFHYKNTTLEDKILKAFNLGLIIFLFLG
REFERENCE	2 (bases 1 to 4888)		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	GTYYICLHNNIIONISIFGRDLIGSDWINGMHTQRMGFPEYSNLIIPITVTVNLYIYI
AUTHORS	Kong, F. and Gilbert, G.L.		/db_xref="GI:18029177"	YMLKRNYSIMTIGVLLFTPLIPIGSGRAGIVAILAQMFLILLNTVVVKKTKIFLL
TITLE	Direct Submission		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	YILPFLAVIMVLYEDNLISLYRIINLRSGSSSEPSFLSKYKDTNIVNINSLPEGV
JOURNAL	Submitted (22-MAR-2001) CIDMLS, ICPMR, Westmead Hospital, Darcy		/db_xref="GI:18029178"	KELMNSDLPLGSHSTYIGYFKSGLGLMNIIVPGLLLIFITNIGRKAQKQSAFYFIEVG
	Road, Sydney, NSW 2145, Australia		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	TLITLFSFALDELDDGANWLLIVFTVLGILENDKDFYSQLKRWKS"
FEATURES	Location/Qualifiers		/db_xref="GI:18029178"	
source	1. 4888		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/organism="Streptococcus agalactiae"		/db_xref="GI:18029178"	
	/mol_type="genomic DNA"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/strain="SG99/276"		/db_xref="GI:18029178"	
	/serotype="Ia"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/db_xref="taxon:1311"		/db_xref="GI:18029178"	
	<1. 135		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
gene	/gene="cpsD"		/db_xref="GI:18029178"	
CDS	/gene="cpsD"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/codon_start=1		/db_xref="GI:18029178"	
	/transl_table=11		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/product="CpsD"		/db_xref="GI:18029178"	
	/protein_id="AAL56388.1"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/db_xref="GI:18029172"		/db_xref="GI:18029178"	
	/translation="AKEQMEQSGSKFLGIILNKVSVSATVGYDGYNGYKRDGRK		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	361. 1536		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
gene	/gene="cpsE"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
CDS	361. 1536		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/gene="cpsE"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/codon_start=1		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/transl_table=11		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/product="CpsE"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/protein_id="AAL56389.1"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/db_xref="GI:18029173"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/translation="MWLKYSPYIIFSSLFIFKNSFTTTLRSLFFSPFIAMNSILLVL		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	LNSFLKYRKYSAKFSRDTKVILITNKSLSKMTFRNKYDHNIVAVCLDSSKDCY		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	DLKNSLRINKDALSELTCITVDQAPINIPILFPGYQIQDIIINDIEAMGVIVNV		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	VEALSFDNIGBKRIQTTEGYSVITYSMKFKYSHLIARPLDITGAILIGLICIVAI		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	FLVPQIRKDGPAIFSONVRGRNRIFRFYKFRSMRVDAEQIKKDLVHNMOTGLMFK		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	LDDPRITKIGKFIKTSIDELPQFYNVLKGDMSLVGTFRPTVDSEYKYNSTOKRRLS		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	FKPGITGLWQISGRNNTDFDEIVKLDVQYINWSIDIKIILLITLKVLLGTGAK"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	1560. 2009		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
gene	/gene="cpsF"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
CDS	1560. 2009		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/gene="cpsF"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/codon_start=1		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/transl_table=11		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/product="CpsF"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/protein_id="AAL56390.1"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/db_xref="GI:18029174"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/translation="MKICLVGSSGGHIAHLNLLKPIWEKDRFWTTPDKDARSILRE		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	BIVHCFPTNRNKNLVKNITLAPKVLKRPDVLISSGAAVAVPEFYIGKLFCKT		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	VYIEVFDRIDPRTGLKLVYPVTDKFIQVWEEMKVPKAINLGIIF"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	2009. 2491		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
gene	/gene="cpsG"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
CDS	2009. 2491		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/gene="cpsG"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/codon_start=1		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/transl_table=11		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/product="CpsG"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/protein_id="AAL56391.1"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/db_xref="GI:18029175"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/translation="WIFVTGTHQEQFNRLIKEVDRLKGTGAIDQEVFIQYGSDFEP		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	361. 1536		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
gene	/gene="cpsH"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
CDS	361. 1536		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/gene="cpsH"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/codon_start=1		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/transl_table=11		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/product="CpsH"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/protein_id="AAL56392.1"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/db_xref="GI:18029176"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/translation="MQLLLLALVLLICSSYNKMKFLNMAEIFFVIFYIYLTSL		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	LHSLFKNPFDRILAFNLSLIGIVAFHYKNTTLEDKILKAFNLGLIIFLFLG		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	GTYYICLHNNIIONISIFGRDLIGSDWINGMHTQRMGFPEYSNLIIPITVTVNLYIYI		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	YMLKRNYSIMTIGVLLFTPLIPIGSGRAGIVAILAQMFLILLNTVVVKKTKIFLL		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	YILPFLAVIMVLYEDNLISLYRIINLRSGSSSEPSFLSKYKDTNIVNINSLPEGV		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	KELMNSDLPLGSHSTYIGYFKSGLGLMNIIVPGLLLIFITNIGRKAQKQSAFYFIEVG		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	TLITLFSFALDELDDGANWLLIVFTVLGILENDKDFYSQLKRWKS"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	3726. 4700		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
gene	/gene="cpsI"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
CDS	3726. 4700		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/gene="cpsI"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/codon_start=1		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/transl_table=11		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/product="CpsI"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/protein_id="AAL56393.1"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/db_xref="GI:18029177"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	TDSNGEICDNLSDQEDNRILVPHKKNQGVSSARNIGLDKSTGEFITFVDSDDFVAPNMI		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	EIMLKNLITENADIADFDISNERDYKKGRRNFYKVKNNNSKEFLSGNRVENIV		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	CTKLYKKSIIIGNLAFDENLKGEDLLFNCKLLCOEHRIVVDVTSLSLYTYRIVKTSAMN		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	QKFNESLDFTITFNEVSSILVPAKLAVKAEKFLREKIKLKRKMFELGNSINDKIKVQ		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	REIFPKDIKSPFPFYKAVKYLKLLSPYLMKCSFKLYVMAYRPFQK"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	4734. 4888		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
gene	/gene="cpsJ"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
CDS	4734. 4888		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/gene="cpsJ"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/codon_start=1		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/transl_table=11		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/product="CpsJ"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/protein_id="AAL56394.1"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/db_xref="GI:18029178"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	/translation="MIKKIEKDLISVIVPIVNVEDYLVECIIESLIVQYRNIEILLIN		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
	DGSTDNCA"		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
ORIGIN			/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
Query Match	10.4%; Score 727.8; DB 1; Length 4888;		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
Best Local Similarity	59.2%; Pred. No. 3.9e-73;		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
Matches 1389; Conservative	0; Mismatches 917; Indels 41; Gaps 7;		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	
Qy	16 TGGCAATTATTTGATAGTAGCAGTTCGCAATTTCTGCAATCTTAAAGTATATACAA		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	75
Db	182 TAGCGATGATTCAACACAGTTGTGTTTATTTTCTGCAAGTTTGCATTAATAATTA		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	241
Qy	76 ATGCTGATTTAAATCG---TTCTGGAATTTTATCATATATGCTGCTTATTTTGCAT		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	132
Db	242 CTCCCAATTTTAAAGCAATAAAGATTATTGTTGTTGTTCTATTGATACATATTATGCT		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	301
Qy	133 TTTTATATCTCGTATGCCAGTTGAAATTTGAGTATAGAGGTAATCTGATAGAGTTGAAA		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	192
Db	302 TTTATCTTCTGATTTTACAGACACTTTTGGAGTCGTGGCTAICTTGAAGAGTTTAAA		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	361
Qy	193 AAACATTAACTAGTATATAATATTGCAATTTTCTTACGGCAGTATCATTTTGTGG		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	252
Db	362 TGGTATTGAAATACAGCTTTTACTATATTTTCTATATCAAGTTTCATTTTATTTTATA		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	421
Qy	253 AGAATAATTTCCGACTTTCAAGACGTGGTCCGCTGTTATTTTACATTAATAAATCGTTT		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	312
Db	422 AAAAATCATTTTACAAACGACACTTTTCTTTTCTTTTCTTTTATGCTATGAATTCGATT		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	481
Qy	313 TGGTATACCTATTTAAACGTAATTTAAGCAGTTTAAAGATAGCTTTCTATTTTTCGACAA		/translation="MQRILVLIPIYNSEAYLKECVQSVLQOHTPLELIEVILDDGS	372



Db 482 TATTGTATCTATTGAAATTCATTTTTTAAATAATTATCGAAATAATTTCTTAGCGTAAGTTTT 541  
Qy 373 TCTATCAAAAAGACAGATTTCTAATTTACACGGCTGACGATGGAAATATGCAAGTTTT 432  
Db 542 CACGAGATACCAAAAGTTGTTTGTATACGAATAAGGATTTCTTTATCAAAAATGACCTTTA 601  
Qy 433 TATTGTAATCACATAAAACAAATTCAAAATAATCTGTGTGATAGTTAGTTAGGTACAG 492  
Db 602 GGAATAAT---ACGACCAATAATTTATTTGCTGTCTGCACTTGGACTCTCTGAAAGG 658  
Qy 493 AAATAGATAAAATTAATTTATCATTTACCGCTCTAATTTCTGTGGAAGAAGCTATAGAGT 552  
Db 659 ATTGTTATGATTGAAACATAAATCGTTTAAGGATAATAAAACAAAGATGCTCTTACTTCAG 718  
Qy 553 TTTCAACAGGGAAGTGGTGCACACAGTCTTTTATAATCTACAAAGTGAGTTTTTAGACG 612  
Db 719 AGTTAACTGCTTAACTGTTGATCAAGCTTTTATTAACATACCCATTTGAAATTAATTTGGTA 778  
Qy 613 TAAAGCAATTC-----GTTTCAGATTTTGAAGTTGTTAGGTAATTTGATGTAAGCGTTG 663  
Db 779 AATACCAATAACAAGATTAATTAATGACATTTGAAGCAATGGGAGTGATTTGTCAATGTTA 838  
Qy 664 ATATTAAATTCATTCGGTTTTTACTCGCTTGA AAAACAAAATAATCCAACTGCTAGTGACC 723  
Db 839 ATGTAGAGGCACCTTAGCTTTTGATAATATAGGAGAAAGCGAATCCAAACTTTTGAAGGAT 898  
Qy 724 ATAGCATTTGAACCTTTTCCAAATTTTATAGCCCTAGTCAATCATGATGAAACGAC 783  
Db 899 ATAGTGTATTACATAATCTATGAATTTCTATAAATATAGTCACTTTATAGCAAAACGAT 958  
Qy 784 TTTTGGATATACTCGGAGCGTAGTCGGTTTAAATTTATTTGGTGTATAGTTTCTAATTTGT 843  
Db 959 TTTTGGATATACCGGGTGTATTAATAGTTTGTCTATATGCGCAATTTGGGCAATTTTTC 1018  
Qy 844 TAGTTCAAATTTATTCGTAGAGATGGTGCACCGGCTATTTTGTCTCAGAAAACGATTCGGAC 903  
Db 1019 TAGTTCCACAAATCAGAAAAGATGGTGCACCGGCTATCTTTCTCAAAATAGAGTAGTGC 1078  
Qy 904 AGAATGGAGCAGATTTTACATCTTACAAAGTTTCGATCGATGATGTTGATGCTGAGGAGC 963  
Db 1079 GTAAATGGTAGGATTTTATAGTTCTATAAATTCAGATCAATGCGAGTAGATGCGAACAAA 1138  
Qy 964 GCAAAAAGACTCTGCAGCCAAACAGATGCAAGGCTGGTATGTTTAAATGGGAA 1023  
Db 1139 TTAAGAAGATTTATTTAGTTCACAATCAAAATGACAGGG-----CTAATGTTTAAAGTTAGA 1193  
Qy 1024 AAACGATCCTAGAAATTTACTCCAATTTGGACATTTTCATACGCCAAAACAAAGTTTACACGAGT 1083  
Db 1194 CGATGATCCTAGAAATTTACTAAATTAGGAAATTTATTTCG-AAAAACAAGCATAGATGAGT 1252  
Qy 1084 TACCACAGTTTATTAATGTTTAAATTTGGCGATATGAGTCTAGTTGGTACAGCTCCACTA 1143  
Db 1253 TGCCTCAATTTCTAATGTTTTTAAAGGTGATATGATGAGTTAGTAGGAACACGCGCTCCCA 1312  
Qy 1144 CAGTTGATGATTTGAAAATAATATACTCTGCTCAAAAGAGACGATTCGAGTTTAAACCGAG 1203  
Db 1313 CAGTTGATGATATGAAAAGTATAATTTCAACGCAAGACGCGCTTAGTTTAAAGCCAG 1372  
Qy 1204 GGATTTACAGGTTCTCTGGCAGGTTAGTGGTCTGATGATATATACACAGACTTCGACGACGTAG 1263  
Db 1373 GAATCACTGGTTTGTGGCAATATCTGGTAGAATAATATACACTGATTTTGTGATGAATCG 1432  
Qy 1264 TTCGGTTGGACTTAGCATAATGATTAATTGGACTATCTGGTGCAGATTTAAATTTTAT 1323  
Db 1433 TAAAGTTAGATGTTCAATATATCAATGAATGCTCTATTGGTCAGATATTAAAGATTATTC 1492  
Qy 1324 TAAAGACAGTCAAAAGTTCTATTGTTGAGAGGGAAGTAAAGTAAAGTA----- 1372  
Db 1493 TCCTAAACACTAAAGGTAGTCTTACTTGGGACAGGTCCTAAGTAAAGGTAAAGTTTGAAG 1552  
Qy 1373 -----TATGAAAGTTTGTGTTGGTTCGTTCTTTCAGGGGGACATTTGACTCACTTGTATTT 1426  
Db 1553 GAATATAATGAAATTTGTTCTGGTTGTTCAAGTGGTGGTCTATCTAGCACACTTTGAACCT 1612

Qy 1427 GTTAAACCGTTTTTGGAGGAAGAAACGTTTTTTGGTGAACATTTTGTATAAAGAGGATGC 1486  
Db 1613 TTTGAAACCCATTTGGGAAAAGAGATAGGTTTTGGTGAACCTTTTGTATAAAGAGATGC 1672  
Qy 1487 AAGAAGTCTTTTGAAGAAATGAAAAAATGTATCCATGTATCTTTCCAAACAAATCGCAATCT 1546  
Db 1673 TAGGAGTATTCTAAGAGNAGAGATTGTATATCATTTGCTCTTTCCAAACNACCGTAATGT 1732  
Qy 1547 CATTAATTTAGTGAATAATACTTTCTTAGTCTTTCAAAATTTTACGTGATGAGAAACGAGA 1606  
Db 1733 CAAAAACTTGGTAAAAAATACTATTCTAGCTTTTAAAGGTCCTTTAGAAAAAGAAACGACAGA 1792  
Qy 1607 TGTATTATTCTCATCTGGTGGCGCGTTCGTCTCCCTCTTTTACATCGGAAACATTT 1666  
Db 1793 TGTATCATATCATCTGGTGGCGCTGTAGCAGTACCATTCTTTTATATTTGTTAGTATTT 1852  
Qy 1667 TGGAGCAAGACGATTTATATCGAAGTATTGTATCGAGTTTAAATAATCTACATTTAACTGG 1726  
Db 1853 TGGTTGTAGACCGTTTTATATAGAGTTTTCGACAGATAGATAAACCACTTTGACAGG 1912  
Qy 1727 AAAACTAGTTTTATCCCGTAAACAGATATTTTTTATTTGTTTCAGTGGGAGAAATGAAGAAGT 1786  
Db 1913 AAAATTAGTGTATCTCTGTAACAGATAAAATTTATTGTTTCAGTGGGAGAAAATGAAAAAGT 1972  
Qy 1787 ATATCCTTAATCTATTAACTTGGGGAGTATTTTTTAATGATTTTGTAAACAGTAGGAACT 1846  
Db 1973 TTATCCTAAGGCAATTTAATTTAGGAGGAATTTTTTAATGATTTTGTCAAGTGGGACCA 2032  
Qy 1847 CATGAACAACAGTTTAAATCGATTGATAAAGAGATGATTATTTGAAAAAAAATGGAAGT 1906  
Db 2033 CATGAACAGCGTTCAACCGCTTTTAAAGAGATTGATAGATTAAAGGGACAGGTGCT 2092  
Qy 1907 ATAAACGAGCAATAATTTATTTCAAAACAGGATATTTCTGACTATATTTCCAGAAATTTGCAAG 1966  
Db 2093 ATTGATCAAGAAGTGTTCATTCAACCGGTTACTCAGACTTTGAAACCTCAGAAATTTGTCAG 2152  
Qy 1967 TATAAAAAATTTCTCAGTTACAAAGAAATGGAACAATATATTAAACAATCAGAAGTAGTT 2026  
Db 2153 TGGTCAAAAATTTCTCTCATATGATGATATGAATCTTTACATGAAGAAGCTGAGATTGTT 2212  
Qy 2027 ATTTGCCAGCGAGCGCCGCTACTTTTATGAAATTCATTATCCAAAGGAAAAACAATTA 2086  
Db 2213 ATCACAATGCGCGTCCAGCGAGTTTATGAAATGCGAGTTTCTAAAGGGAAAAACAATTT 2272  
Qy 2087 TTGTTTCTTAGACAAAAAAGTATGTTGAACATGTAAATGATCATCAAGTAGAGTTGTA 2146  
Db 2273 GTGGTTCTTAGACAAGAAACAGTTTGGAGAGCATGTGAATAATCATCAGGTGGATTTTGTG 2332  
Qy 2147 AGAAGAAATTTTACAGATATAATAATTTTATTTATAGAAAATATAGATTTGTTGAA 2206  
Db 2333 AAGGAGTTATTCTTTGAAATATGAATTAGATTATATTTTGAAATATCAGTGAATTAGAGAT 2392  
Qy 2207 AAAATTATTGAGTTTCTAAGCAAC---TAACTTTACATCAATAATAATTTTTTTTGT 2263  
Db 2393 ATTATTAAAGAAAAAATAATATCTACTAGTAAAGTAAATATACAAACATGATTTTGT 2452  
Qy 2264 GAAAGATTAAAACAAAATAGTTGAAAAATTTTAAATGAGGATCAAGAAAAATGAATAATAAAA 2333  
Db 2453 TCCTCTTTTCAAAAATGAACCTTCTAAACTATTTTGAATAAATAATATTTTGTGGAGAAAAA 2512  
Qy 2324 AGATGCA 2330  
Db 2513 AATTGAA 2519

RESULT 20  
BD166249  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS

BD166249 6865 bp DNA linear PAT 17-JAN-2003  
Beta-1,3-galactosyltransferase and DNA encoding the enzyme.  
BD166249  
BD166249.1 GI:27872061  
JP 2002199885-A/2.

SOURCE unidentified  
ORGANISM unidentified  
REFERENCE unclassified.  
1 (bases 1 to 6865)  
AUTHORS Miyake, K., Watanabe, M. and Iijima, S.  
TITLE Beta-1,3-galactosyltransferase and DNA encoding the enzyme  
JOURNAL Patent: JP 2002199885-A 2 16-JUL-2002;  
KYOWA HAKKO KOGYO CO LTD  
COMMENT OS Streptococcus agalactiae Type Ib  
PN JP 2002199885-A/2  
PD 16-JUL-2002  
PF 05-JAN-2001 JP 2001000392  
PI KATSURIBE MIYAKE, MASAKI WATANABE, SHINJI IJIJIMA PC  
C12N15/09, C12N19/21, C12N9/10, C12P19/28// (C12N15/09, C12R1:46), PC  
(C12N9/10, C12R1:19), (C12P19/28, C12R1:19), C12N15/00, (C12N15/00, PC  
C12R1:46)  
CC Beta-1,3-galactosyltransferase and DNA encoding the enzyme FH  
Key Location/Qualifiers  
FT CDS (617).. (1789)  
FT CDS (1816).. (2262)  
FT CDS (2265).. (2744)  
FT CDS (2843).. (3979)  
FT CDS (3982).. (4953)  
FT CDS (5009).. (5947).  
FEATURES  
source  
1..6865  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"  
ORIGIN  
Query Match 10.4%; Score 727.8; DB 6; Length 6865;  
Best Local Similarity 59.2%; Pred. No. 3.5e-73;  
Matches 1389; Conservative 0; Mismatches 917; Indels 41; Gaps 7;  
Qy 16 TGGCAATTAATGATAGATAGCAGTTGCAATTTCTGCAATCTTAAGTCAATATACCAA 75  
Db 438 TAGCGATGATTCAAACCGTTGGTATTATTTCTGCAAGTTTGACATTAACATTAATA 497  
Qy 76 ATGCTGATTAATCG---TTCTGGAATTTTATCATATGATGTTTCATTTATTTGCAAT 132  
Db 498 CTCCTCAATTTAAAGCAATTAAGATTTATTGTTGTTGTTCTATGATATATTTGTTCT 557  
Qy 133 TTTTATATCTGATGCGAGTTGAAATTTGATATAGAGTAATCTGATAGAGTTTGAAA 192  
Db 558 TTTATCTTCTGATTTTACAGAGACTTTTGAGTCTGGCTAUCTTGAAGTTTAAA 617  
Qy 193 AAACATTTAATAGTATAATTTGCAATTTTCTTACGCGAGTATCATTTTGTGG 252  
Db 618 TGGTATTGAATATACAGCTTTTACTATATTTTCTATATCAAGTTTCATTTTATTCTA 677  
Qy 253 AGAATAATTTGCACTTTCAAGACGTTGGTGGCTGATTTTCATATTAATTAATCTGTTT 312  
Db 678 AAAACTCTTTTACACGACGACTTTCTCTTTTCTTTTATTTGCTATGAAATTCGATTT 737  
Qy 313 TGGTATACCTATTAAACGTAATTTAAGACGTTTAAAGATAGCTTTCTATTTTCGACAA 372  
Db 738 TATATATCTATGAAATTTCAATTTTAAATATATATCGAAATATTTCTTACGCTAAGTTT 797  
Qy 373 TCTATCAAAAAGAACGATTTCTAATTTACAAACGCTTGAACGATGGGAAATATGCAAGTTT 432  
Db 798 CACGAGATACCAAGTTGTTTGTGATAACGAATAAGGATTTCTTATCAAAAATGACCTTTA 857  
Qy 433 TATTTGAATACATAAAACAAATTCAAAATCTTTGTTGCAATGTTGATGTTTATGATACAG 492  
Db 858 GGAATAAAT---ACGACCATTAATATATCGCTGCTGATCTGATCTTGGATCTCTGAAAAG 914  
Qy 493 AAATAGATATAAATTAATTTATCATTTACCGCTCTATTTTCTGTGGAAGAGCTATAGAGT 552  
Db 915 ATTGTTATGATTTGAAAACATAACTCGTTAAGGATAATAAAACAAAGATGCTCTTACTCAG 974  
Qy 553 TTTCAACAAGGAAGTGTGCGACGCTCTTTTATAAATCTACCAAGTGGATTTTATAGACG 612

Db 975 AGTTAACTGCTTAACCTGTTGATCAAGCTTTTATTAACATATCCCATTAATTTGGTA 1034  
Qy 613 TAAAGCAATTC-----GTTTCAGATTTTGGATTTCTTAGGTATTCATGTAAAGCGTTG 663  
Db 1035 AATACCAATACAGATATTAATGACATTTGAAGCAATGGAGTGTATGTCATGTTA 1094  
Qy 664 ATATTAAATTCATTCGGTTTACTGCGTTGAAAAACAAAAAATCCAACCTGCTAGTGACC 723  
Db 1095 ATGTAGAGGCACTTAGCTTTGATATATAGAGAAAAAGCAATCCAAATCTTTTGAAGAT 1154  
Qy 724 ATAGCATTTGTAACCTTTTCCACAAATTTTATAAGCCTAGTCATATCATGATGAACACG 783  
Db 1155 ATAGTGTATTACATATTTCTATGAATTTCTATAAATATAGTCACCTTATAGCAAAACGAT 1214  
Qy 784 TTTTGGATATATCTCGGAGCGGTAGTCGGGTTAAATTTATTTTGGTATAGTCTTCTATTTGT 843  
Db 1215 TTTTGGATATCATGGTGTATTATAGGTTTGTCTCATATGTGGCATTGTGGCAATTTTC 1274  
Qy 844 TAGTTCCAAATTTATTCGTAGAGATGGTGACCGGCTATTTTGTCTCAGAAAACGATTTGAC 903  
Db 1275 TAGTTCCGCAATTCAGAAAAGATGGTGACCGGCTATCTTTTCTCAAAATAGAGTAGGTC 1334  
Qy 904 AGAATGGACGCAATTTTACATTTCTACAGTTTCGATCGATGATGATGCTGCTGAGGAC 963  
Db 1335 GTAAATGGTAGGATTTTATAGATTTTATAAATTCAGATCAATCGAGTAGATGCGAACAAA 1394  
Qy 964 GCAAAAAGACTGTCTCAGCCAAAACAGATGCAAGGGTGGGTATGTTTTAAAAATGGGAA 1023  
Db 1395 TTAAGAAAGATTTATTAGTTCACATCA-----ATGACGGGGCTAATGTTTAAAGTTAGA 1449  
Qy 1024 AAACGATCTAGAAATTAATCTCAATTTGGACATTTTATACGCAAAAACAAGTTTACAGAGT 1083  
Db 1450 CGATGATCTAGAAATTTACTAAAATAGGAAAATTTATTCTG-AAAAACAAGCATAGATGAGT 1508  
Qy 1084 TACCACAGTTTATAAGTTTAAATTTGCGGATGATGAGTCTAGTTGGTACAGCTCCACCTA 1143  
Db 1509 TGCCTCAATTTCTAATGTTTAAAGGTGATATGAGTTTATAGTAAAGACACGCGCTCCCA 1568  
Qy 1144 CAGTTGATGAATTTGAAAAATATATCTCTGGTCAAAAGAGAGCAATGAGTTTAAACACAG 1203  
Db 1569 CAGTTGATGAATATGAAAAGTATTAATTCACGCAAGACGCGCTTAGTTTAAAGCCAG 1628  
Qy 1204 GGATTACAGGTCCTCGGCAAGTTAGTGGTCTAGTAAATATCAAGATTCGACGACGCTAG 1263  
Db 1629 GAATCAGCTGTTTGTGGCAATATCTGTAAGAAATATATTTACTGATTTGATGAAATCG 1688  
Qy 1264 TCGGTTGGACCTTAGCATACATTTGATTAATTCGACTATCTGGTTCAGATATTAATTTAT 1323  
Db 1689 TAAAGTTAGATGTTCAATATATCAATGAATGGTCTTATTGGTTCAGATATTAAGATTTTC 1748  
Qy 1324 TAAAGACAGTCAAAAGTTGCTATTGTTTGAGAGAGGGAAGTAAGTAAAGTA----- 1372  
Db 1749 TCCTAACGCTAAAGTAGTTTACTCGGGAAGAGGCTAAGTAAAGGTAAGTTTGAAG 1808  
Qy 1373 -----TATGAAAGTTTGTGTTGGTGGTCTTCTAGGGGGGACATTTGACTCTGTTATTT 1426  
Db 1809 GAATATAATGAATTTGTTCTGGTTGGTTCAAGTGGTGTCTACCTAGCACACTTGAACCT 1868  
Qy 1427 GTTAAACCGTTTGGAAAGAAAGAACGTTTGGTGTAAACATTTGATTAAGAGATGTC 1486  
Db 1869 TTTGAAACCCATTTGGGAAAAAGAGATAGGTTTGGGTAACTTTTGGTAAAGAGATGTC 1928  
Qy 1487 AAGAAGCTTTTGAAGATGAAAAATGATTCATCTTCCATGTTTACCTTCAACAAATCGCAATCT 1546  
Db 1929 TAGGAGTATTTCTAAGAGAGAGATTTGTATATCATTTGCTTCTTCCAAACAAACCGTAATGT 1988  
Qy 1547 CATTAAATTTAGTGAATAATCTTTCTTAGCTTTTCAAAATTTTACGTGATGAGAAACGAGA 1606  
Db 1989 CAAAACTTGTGTAATAATACTATTTCTAGCTTTTAAAGTCTCTAGAAAGAAAGACGAGA 2048  
Qy 1607 TGTTATTTATTTCACTGCTGGCGGCTGCTGCTCCCTTCTTTTATCATCGGAAAACTATT 1666  
Db 2049 TGTATCATATCATCTGCTGCGCTGTAGCAGTACCACTTCTTTTATATTGTTGTAAGTTATT 2108

Qy	1667	TGGAGCAAGACGATTTATATTGAAGTATTTGATCGAGTTAATAAATCTACATTAAGTGG	1726
Db	2109	TGGGTGAAGACCGTTTATATAGAGGTTTTCGACAGGATAGATAAACAATTTTGACAGG	2168
Qy	1727	AAAACTAGTTTATCCCGTAACAGATATTTTATTGTTTCAGTGGGAAGAAATGAAGAAGGT	1786
Db	2169	AAATATTAGTGATCTGTAAACAGATAATTTATTGTTTCAGTGGGAGAGAAATGAAGAAAGT	2228
Qy	1787	ATATCCTAAATCTATTAATCTTGGGAGATTTTATTGATTTTAAATGATTTTGTAAACAGTAGGAAT	1846
Db	2229	TTATCCTAAGGCAATTAATTTAGCAGGAATTTTAAATGATTTTGTACAGTAGGAGACA	2288
Qy	1847	CATGAACAACAGTTTATTCGATTTGATAAACAGATTTGATTTTGAAGAAAAATGGAAGT	1906
Db	2289	CATGAACAACAGTTTCAACCGCTTATTAAAGAAGTTGATAGATTAAAGGGACAGGTGCT	2348
Qy	1907	ATAACCGACGAAATATTATTCAACAGGATATTCTGACTATATTCCAGATAATTGCAAG	1966
Db	2349	ATTGATCAAGAAGTGTTCAATTCAAACGGTTACTCAGACTTTGAACCTCAGAAATGTCAG	2408
Qy	1967	TATAAAAATTTCTCAGTTACAAAGAAATGGAACAATATATTAACAATCAGAAGTAGTT	2026
Db	2409	TGGTCAAAATTTCTCATATGATGATGAATGAATCTTACATGAAGAAGCTGAGATTGTT	2468
Qy	2027	ATTTGCCACGAGGCCCGCTACTTTTATGAATTCATATTCATCAAGGAAAAACAATTA	2086
Db	2469	ATCACACACGCGGTCGACGAACGTTTATGAATCGCGTTTCTAAAGGGAACAACTATT	2528
Qy	2087	TGTGTTCTCAGACAAAATAAGTATGGTGAACATGTAAATGATCAATCAAGTAGAGTTGTA	2146
Db	2529	GTGGTTCTAGACGAAGACAGTTTGGAGAGCATGTGAATATCAATCAGGTGGATTTTGT	2588
Qy	2147	AGAAGAATTTTCAAGATAATAATATTATTTATTTATAGAAATATAGATGATTTGTTCAA	2206
Db	2589	AAAGAGTTATTCTTGAATATAGTTAGATTATATTTTGAATATCAGTGAATTAGAGAAAT	2648
Qy	2207	AAATATTAGTAAATTTCTAAGCAAC----TAACATTACATCAATATAATTTTTTTTGT	2263
Db	2649	ATTATTAGGAAAAAATATATCTACTAGTAAAGTAATATCACAACAAATGATTTTTGT	2708
Qy	2264	GAAGATTAAACAAATAGTTGAAATAATTTAATGAGGATCAAGAAATGAATAATAAAAA	2323
Db	2709	TCCTCTTTCAAAATGAACCTTTCTAACTATTTGAATAATAATATTTTGTGGAGAAAA	2768
Qy	2324	AGATGCA 2330	
Db	2769	AATTGAA 2775	

RESULT 21	
AB050723	
LOCUS	AB050723 9987 bp DNA linear BCT 03-FEB-2001
DEFINITION	Streptococcus agalactiae cpsIBD, cpsIBF, cpsIBG, cpsIBH, cpsIBI, cpsIBJ, cpsIBK, cpsIBL, neuB, neuC genes, complete cds.
ACCESSION	AB050723
VERSION	AB050723.1 GI:12697593
KEYWORDS	
SOURCE	Streptococcus agalactiae
ORGANISM	Streptococcus agalactiae
REFERENCE	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
AUTHORS	1. (sites)
TITLE	Miyake, K., Watanabe, M. and Iijima, S.
JOURNAL	CpsJ of Streptococcus agalactiae type 1b shows beta-1,3-galactosyltransferase activity
AUTHORS	Unpublished
TITLE	2. (bases 1 to 9987)
JOURNAL	Miyake, K., Watanabe, M. and Iijima, S.
AUTHORS	Direct Submission
TITLE	Submitted (01-NOV-2000) Katsuhide Miyake, Nagoya University, Grad. Sch. of Engineering, Dept. of Biotechnology; Chikusa-ku, Furo-cho, Nagoya, Aichi 464-8603, Japan

FEATURES	
source	1. .9987
Location/Qualifiers	
/organism="Streptococcus agalactiae"	
/mol_type="genomic DNA"	
/db_xref="taxon:1311"	
425..430	
442..1791	
/gene="cpsIBD"	
/gene="cpsIBD"	
/function="glucosyltransferase"	
/codon_start=1	
/evidence=not_experimental	
/transl_table=11	
/product="CpsIBD"	
/protein_id="BAB21597.1"	
/db_xref="GI:12697594"	
/translation="MIQTVVYFSASLTLTITPNFKSNKDLLFVLLIHIVFVLSDFYRDFWSRGYLBEEFMVKLYSFYIFISSLFFISKNSFTTTRLSFFFIAMNSILLYL LNSFLKYRYKYSARFSDTKVVLITNKDSLKMTFRNKYDHNYYIACVILSSSKDCY DLKHSRLRIINKDALTSCLTVDQAFINIPTELFGYQIQDIINDIEMGVIVNVN FLVSLPDNIGEKRIQTPEGYSVITYSMKFYKSHLIAKRFIDIMGALIGLLICGIVAI FLVQIRKDGGAIPFSONRVGRNGRIFRFYKFSRMRVDAQIKDILVHNMOTGLMEK LDDPRITKIKGFIKRTSIDELPQFTNVLAGDMSLVGTREPTVDEYEKYNSTQKRRLS FKPGITGLWQISGRNNITDFDEIVKLDQVQINWSIWSDIKILLTLKVVLLGTGAK"	
1803..1810	
/gene="cpsIBF"	
/db_xref="GI:12697595"	
/translation="MKICLVSSGGHLAHLNLLKPIWEKEDRFWVTFDKEDARSILREI VYHGCFPTNRNVNKLAVKNTILAFKVLKRPDPVITSSGAAVAPFPFYIGKLFCKT VYIEVDFRIDKPTLTGKLVYPTVDKFIQWEEMKKVYPKAINLGGIF"	
2249..2256	
/gene="cpsIBF"	
2284..2746	
/gene="cpsIBG"	
2264..2746	
/gene="cpsIBG"	
/function="beta-1,4-galactosyltransferase"	
/codon_start=1	
/evidence=not_experimental	
/transl_table=11	
/product="CpsIBG"	
/protein_id="BAB21599.1"	
/db_xref="GI:12697596"	
/translation="MIFVTYGTHEQQPNRLIKEVDRLKGTGAIDQEVFIQTGYSDPEP QNCWSKFLSDYDMNSTKREIVITHGGPATFMNAVSKGKTIVFPRQGFQGHVNN HQVDFLKLFLKYELDYILNISLENIIEKNISTSKVISQNNDFCSSFKNLSKLIFE"	
2806..2812	
2842..3981	
/gene="cpsIBH"	
2842..3981	
/gene="cpsIBH"	
/function="putative CP polymerase"	
/codon_start=1	
/evidence=not_experimental	
/transl_table=11	
/product="CpsIBH"	
/protein_id="BAB21600.1"	
/db_xref="GI:12697597"	
/translation="MOLLILALIVLLICSSYNEKMKFLNMAEIFFIVFVNVILVSI LNSLFRSPFHRVIAAFNSLAVGSVLLFYHYKNTNIELTKLKSFLFNAILPCLG FLYYALYFDVENVSUGFNLIGSDWMGHMTQRAMAFEYSNLIIPLTITNIYIYI	

```
YIKORYSSGMILGALLSTIILPIGSGSRAGIIVVLQVLIILLNTIVIKQOTIRPFL
YLVPLILLVLPFDNLUVSINRILNRSSESRFLYKDTVHVSVITLDFLFGKV
KELMNLSDHSTYIGYFYKTLGLINVLGLFLILISIIKEAKSDFYBEING
SVILFSPFALIEDIGANWLIIFVFTVLGILENDKFYSQAKRWES"
RBS
3964..3970
/gene="cps1bh"
3981..4955
/gene="cps1bi"
3981..4955
/gene="cps1bi"
/function="beta-1,3-N-acetylglucosaminyltransferase"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="Cps1bi"
/db_xref="GI:12697598"
/translation="MEKQLVSIPIVNSEAYLKECVQSVLQOOTHSLIEVILINDGS
TDNSEICDNLQKDDRLIVFHKKNGGVSSARNLGLDKSTGCEFTFVDSDDFVAPNII
EIMLNLTEDADIAVDVFDISNERDYKRRNFYKVKNNLSKFLSGNRVENIV
CTLYKXSIIIGNRFDENLKGEDLLFNCKILCOEHCIIVDTTSSLYTIRIVKTSAMN
QEPNENSLDFTITFEISSIIPAKLANVVEAKFLREKVKCLRMFELGSLGNSIDSKIKLQ
REIFPKDVKLPFYKAVKYLKGLLSIYLMKCSPILYIKLYDRFQK"
RBS
4995..5001
5008..5949
/gene="cps1bj"
5008..5949
/gene="cps1bj"
/function="beta-1,3-galactosyltransferase"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="Cps1bj"
/protein_id="BAB21602.1"
/db_xref="GI:12697599"
/translation="MNRGIIMNSVNEPLNYRDSVESILNQLTDFEPIIIVDNPGRG
DLKQLTYSVDNRKIKILLMEINIGLASSLNKAVKISGKEYIFRMDADDISLPSRSG
KQIFRMEISLDFATIELIDQGNLVYKQRESNKIYLTNDIRKMLNRSILAHPTW
CVKKVFDKMGYDLVDEYDPAIRGALADFKILNKLVLLOVRLNENGISOTNKF
KQIYLSAILODFYKEKSIDITKITNYFOEVVIKRVTQOELSFKVFLKSTPSITIRK
LYCLLYFKSPVLRLLINDINILVLKFGQKQSD"
RBS
5928..5935
5942..6880
/gene="cps1bk"
5942..6880
/gene="cps1bk"
/function="putative sialyltransferase"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="Cps1bk"
/protein_id="BAB21603.1"
/db_xref="GI:12697600"
/translation="MTRKIVYCHTLVHLHLCYKKEIYSNLEIILSSSIDVDNLEK
KLKSTINIHILSESSGEBLLSVLKDAGLSYKFSNCFIPNDATPIGRLLIKHGI
YNNLEDGLNFTYISPSQKLKVVYKVIILHKIQPHGFSRYCLIGIENSLVNLPKDP
RYKFLIEPRKELFDNVEYOKEMAINLFGAVRVSISPSVLITQPLSIDKEPMSYN
NKIETSEQPFYKSIYNEVINKGNVYLRPRDVEDVYSKLPVELLPSPVPMELIEL
MLTGRFEGGITHSTALDFLCVDRKKITLVLDKDIK"
RBS
6865..6870
6877..8277
/gene="cps1bl"
6877..8277
/gene="cps1bl"
/function="putative repeat unit transporter"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="Cps1bl"
/protein_id="BAB21604.1"
/db_xref="GI:12697601"
```

```
/translation="MNTKQLQSGFIYITIGNLVQGLAFITLPIYTRVISTEVYGOYS
LYVAMNIIIMFLIGLQTSLSARVYKGEFKSYGSFSAFVGNMFWLIVLIIAPLFR
SFLAPLVGFSESIYLLWVCOSYANYVVTFFQYIQOORSLANIILSLANAVSVALS
LFLTFHWSDDFLSRVLGVFVFTIITGVAFAYIVHKSFSYFNPKYFRFIVTVSIPLI
HLGHOLLNOLDRLMLARLYSTKEVAMYSGVSLMIITQIVLNSINMAWIWPFEDARK
EKLOLSTYSISRYLXGLVFLTLGYLTPPELAQIMGGDOYSSVQFISLIIVSPLVP
LYTFVNIQFYANTWNPITGLAAGVNNLNLVLIPHAYAGAMATIIISYLALLI
FHHIVSKYKHYSVSVRYIILSGIVFSYAMLNMFLDNIVIRWSLUGIIILIVYSIV
FOKVLIDLLSKRRRR"
RBS
8262..8269
/gene="cps1bl"
8277..9302
/gene="neuB"
8277..9302
/gene="neuB"
/function="sialic acid synthetase"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="NeuB"
/db_xref="GI:12697602"
/translation="MVVILAEIGCNHNDINLAKMVDVAVSCGVDAVKFOTFEKAEKL
ISFPAPKAEYQKATCTGADSOLETKLELSPEYLEMRDVAISGVETFTSTPDEES
LEFLIDTMEPIYKIPSGEITNLPYLEKIQKQKVIILSTGMVWEEIHQAVNLKQNG
TTDISILHCTTEYETPYPSNLNVIHTLKDFKDLTIGYSDHSISEVPIIAAAMGAE
VIEKHFTLDNMEVDPDHKASATPDILAAVLKGFALLNQLALGRFEKIPDPVEEKNI
RKSVALPKIKGDIYSIENTIVKRPNGISPMNWDILGQEAQDDFEDEVIRDSRF
ENOLPEL"
RBS
9367..9374
9379..9987
/gene="neuC"
9379..9987
Query Match 10.4%; Score 727.8; DB 1; Length 9987;
Best Local Similarity 59.2%; Pred. No. 3.1e-73; Indels 41; Gaps 7;
Matches 1389; Conservative 0; Mismatches 91;
Qy 16 TGGCATTATTGATGATAGTAGCAGTTGCAATTTCTGCAATCTTAACAAGTCATATACCAA 75
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
437 TAGCGATGATTCAAACCGTTGTGTTATTCTTCTGCAAGTTTGACATTAACTAATTA 496
Qy 76 ATGCTGATTTAAATCG---TTCTGGAATTTTATCATAAATGATGTTTCATTATTTTGCAT 132
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
497 CTCCCAACTTTAAAGCAATAAAGATTATTGTTTGTCTTATTTGATATGATATATTGTCT 556
Qy 133 TTTTATATCTCGTATGCCAGTTGAAATTGATAGATAGAGTAATCTGATAGATTTGAAA 192
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
557 TTTATCTTCTGATTTTACAGAGACTTTTCGAGCTCGTGGCTATCTTGAAGAGATTTAAA 616
Qy 193 AAACATTAACTATAGTATATAATTGCAATTTTCTTACGGCAGTATCATTTTGTGTTGG 252
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
617 TGGTATTGAATACAGCTTTTACTATATTTTCATATCAAGTTTCATTATTTTATTCTTA 676
Qy 253 AGAATAATTCGCACTTTCAAGACGTGGTGGCGGTGATTTTTCACATTAATAAATTCGTTT 312
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
677 AAAACTCTTTTACAACGACACAGCTTCTCTCTTTTACTTTTTTATTGCTATGAATCGATT 736
Qy 313 TGGTATACCTATTAAAGTAATTAATTAAGCAGTTTAAGGATAGCTTTCTATTTTCGACAA 372
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
737 TATTATCTATCTGAATTCATTTTAAATAATTATCGAAAAATATTCTTACGCTAAGTTT 796
Qy 373 TCTATCAAAAAAGACGATTTCTAATTACACGGCTGACGATGGGAAATATGCAAGTTT 432
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
797 CACGAGATACCAAAGTTGTTTGTATACGAATAAGGATTTCTTATCAAAAATGACCTTAA 856
Qy 433 TATTTGAATCACATAAACAATTCAAAAATCTTGTGTGCAATTTGGTAGTATTTAGGTACAG 492
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
857 GGAATAAAT---ACGACCATTAATTATCGCTGCTGTATCTTGGATCTCTCTGAAAAGG 913
Qy 493 AAATAGATAAATAATTTATCATTAACCGCTCTATTATTCTGTGGGAAGAAGCTATAGAGT 552
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
914 ATTGTTATGATTTGAAACATAACTCGTTAAGGATAATAAACAAGATGCTCTTACTTCAG 973
```

553 TTTTCAACAGGGAAGTGGTCGACCAAGCTCTTTATATAATCTACCAAGTGGATTTTACAGC 612  
Db |||||  
974 AGTTAACTCGCTTAACCTGTTGATCAAGCTTTTATTAACATACCACTGGAATTTTGGTA 1033  
Qy |||||  
613 TAAAGCAATTC-----GTTTCAGATTTGAGTTGTTAGGTATTGATGTAAGCGTTG 663  
Db |||||  
1034 AATACCAATACAGATATTTATTAATGACATTTGAACCAATGGGAGTGAATGTCATGTTA 1093  
Qy |||||  
664 ATATTAATTCATCGGTTTACTCGGTGAAAAACAAAAATCCAACTGCTAGGTGACC 723  
Db |||||  
1094 ATGTAGAGGCACTTAGCTTTGATATAATAGGAGAAAGCGAATCCAACTTTTGAAGCAT 1153  
Qy |||||  
724 ATACCATTTGAACCTTTTCCAAATTTTATAGCCCTAGTGCATATCATGATGAACGAC 783  
Db |||||  
1154 ATAGTGTATTACATATTTCTATGAATTTCTATNAATATAGTACCTTTATAGCAAAACGAT 1213  
Qy |||||  
784 TTTTGGATATACTCGAGCGGTAGTCGGGTTTAAATTTTGGGTATAGTTTCTATTTGT 843  
Db |||||  
1214 TTTTGGATATCATGGGTGCTATTATAGTTTGTCTATATGGGATTTGGCAATTTTTC 1273  
Qy |||||  
844 TAGTTCCAAATTTTCTAGAGATGGTGCACCGGCTATTTTGTCTCAGAAAACGAGTTGGAC 903  
Db |||||  
1274 TAGTTCCGCAATCAGAAAAAGATGGTGCACCGGCTATCTTTCTCAAAATAGAGTAGTGC 1333  
Qy |||||  
904 AGAATGGACGATATTACATTTCTACAAAGTTTCATCGATGATGTTGTGCTGAGGAGC 963  
Db |||||  
1334 GTAAATGGTAGGATTTTATAGATTTCTATAAATTCAGATCAATCGCAGTAGATGCAGAACAA 1393  
Qy |||||  
964 GCAAAAAGACTCTCTCAGCCAAACACAGATGCAAGGGTGGTATGTTTAAATGGGAA 1023  
Db |||||  
1394 TTAAGAAAGATTTATTTAGTTCACATCA-----AATGACGGGGCTAATGTTTAAAGTTAGA 1448  
Qy |||||  
1024 AAACGATCCTAGAAATTTACTCCAATTTGGACATTTTCATACGCCAAACAAAGTTTACAGAGT 1083  
Db |||||  
1449 CGATGATCCTAGAAATTTACTAAATTAGGAAATTTATTTCG-AAAAACAAGCATAGATGAGT 1507  
Qy |||||  
1084 TACCACAGTTTATTAATGTTTAAATTTGCGGATATGAGTCTAGTTGGTACAGCTCCACTA 1143  
Db |||||  
1508 TGCCTCAATTTCTAATGTTTTTAAAGGTGATATGAGTTTAGTAGGAAACACGCCCTCCCA 1567  
Qy |||||  
1144 CAGTTGATGAATTTGAAAAATATACTCTGCTCAAAAGACAGATTCGATTTTAAACACAG 1203  
Db |||||  
1568 CAGTTGATGAATTTGAAAGATTAATTTCAACGCAAGAGCGACGCTTAGTTTAAAGCCAG 1627  
Qy |||||  
1204 GGATTACAGGTTCTCTGGCAGGTTAGTGGTCTGATGATATATACAGACTTCCAGCAGCTAG 1263  
Db |||||  
1628 GAATCACTGGTTTCTGGCAATATCTGTTAGAAATATATTTACTGATTTTGTGAAATCG 1687  
Qy |||||  
1264 TTCCGTTGGACTTAGCATACATTTGATTAATTTGGACTATCTGGTACAGATTTAAATTTTAT 1323  
Db |||||  
1688 TAAAGTTAGATGTTTCAATATATCAATGAATGGTCTATTTGGTACAGATATTAAGATTATTC 1747  
Qy |||||  
1324 TAAAGACAGTCGAAAGTTGTTATTTGTTGAGAGAGGGAAGTAAAGTAAAGTA----- 1372  
Db |||||  
1748 TCCTAAGCTTAAGGTAGTTTACTTCGGGACAGAGCTAAGTAAAGTTAGGTTTGAAG 1807  
Qy |||||  
1373 -----TATGAAAGTTTGTGGTTCGTTCTTCAGGGGACATTTGACTCAGTTGTTATTT 1426  
Db |||||  
1808 GAATATAATGAAATTTGTTCTGGTTGGTTCAAGTGGTGGTCACTAGCACACTTGAACCT 1867  
Qy |||||  
1427 GTTAAACCGTTTGGAGGAGGAAAGAACGTTTGGTGGTAAACATTTGATGAAGAGGATGC 1486  
Db |||||  
1868 TTTGAAACCCCATTTTGGGAAAAAGAGATAGGTTTGGGTAACTTTTGTATGAAGAAGATGC 1927  
Qy |||||  
1487 AAGAAAGTCTTTTGAAGATGAAAAATGTATCCATGTTACTTCCCAAAATCCCAATCT 1546  
Db |||||  
1928 TAGAGATTTCTAAGAGAGAGATTTGATATCATTTGCTTTCTTCCAAACAAACCGTAAATGT 1987  
Qy |||||  
1547 CATTAATTTAGTGAATAATCTTTCTTAGCTTTTCAAAATTTTACGTGATGAGAAACAGGA 1606  
Db |||||  
1988 CAAAAAAGCTTGGTAAAAAATACTATCTTAGCTTTTAAAGTCTCTTAGAAGAAAGAACACAGGA 2047

Qy 1607 TGTATTATTTTTCATCTGGTGGCGGTTGCTGTGCCCTCTTTTACATCGGAAACTATT 1666  
Db |||||  
2048 TGTATCATATCATCTGGTGGCGCTGTAGCAGTACCAATCTCTTTATATTGTAAGTTATT 2107  
Qy |||||  
1667 TGGCAACAAAGCAGATTTATATTGAAGTATTGTATCGAGTTTAAATAAACTCATATTAACTGG 1726  
Db |||||  
2108 TGGCTGTAAAGCCGTTTATATAGAGTTTTCGACAGATAGATATAACCACTTTTGACAGG 2167  
Qy |||||  
1727 AAACTAGTTTATCCCGTAAACAGATATTTTATTTGTTTCAGTGGGAGAAATGAAGAAGGT 1786  
Db |||||  
2168 AAAATTAGTGTATCTCTGTAACAGATAAAATTTATTTGTTTCAGTGGGAGAAATGAAAAAGT 2227  
Qy |||||  
1787 ATATCCTTAATCTTATTAACCTTGGGAGTATTTTATGATTTTGTAAACAGTAGGAACT 1846  
Db |||||  
2228 TTATCCTTAAGCAATTAATTTAGGAGGAATTTTATGATTTTGTGCACATGAGGACA 2287  
Qy |||||  
1847 CATGAACAACAGTTTAAATCGATTGATAAAAGAGATTGATTATTTGAAAAAATAATGGAAGT 1906  
Db |||||  
2288 CATGAACAGAGTTCAACCGTCTTATTAAGAAGTTGATAGATTAAAGGGACAGGTGCT 2347  
Qy |||||  
1907 ATAAACCGACGAAATATTTATTTCAAAACAGGATATTTCTGACTATATTTCCAGAAATTTGCAAG 1966  
Db |||||  
2348 ATTGATCAAGAGTGTTCATTCAAAACGGGTTACTCAGACTTTTGAACCTCAGAAATTTGTCAG 2407  
Qy |||||  
1967 TATAAAAAATTTCTCAGTTACAAAGAAATGGAACATATATTTAAACAATCAGAAGTAGTT 2026  
Db |||||  
2408 TGGTCAAAATTTCTCTCATATGATGATATGAACCTCTTTACATGAAGAAGCTGAGATTGTT 2467  
Qy |||||  
2027 ATTTGCCACGAGCGCCGCTACTTTTATGCAATTTCAATTTCCAAAGGAAAAACAATTA 2086  
Db |||||  
2468 ATCACAACAGCGCGGTCCAGCAACGTTTATGAAATGCGAGTTCTTAAAGGAAAAAATCTATT 2527  
Qy |||||  
2087 TTGTTTCTTAGACAAAAAAGTATGGTGAACATGTAAATGATCATCAAGTAGAGTTTGT 2146  
Db |||||  
2528 GTGGTTCTTAGACAAGAACAGTTTGGAGAGCATGTGAATAATCATCAGGTGGAATTTTGT 2587  
Qy |||||  
2147 AGAAGAAATTTTACAGATAATAATTTATTTATTTATAGAAAATATAGATGATTTGTTGAA 2206  
Db |||||  
2588 AAAAGAGTTATTCTTGAATAATGATGATTATATTTTGAATATCAGTGAATTAGAGANT 2647  
Qy |||||  
2207 AAAATATTCAAGCTTTCTAAGCAAAC---TAACTTTACATCAATAATAATTTTGT 2263  
Db |||||  
2648 ATTATTAAAGAAAAAATATATCTACTAGTAAAGTAAATATACAAACAATGATTTTGT 2707  
Qy |||||  
2264 GAAAGATTAAAACAAAATAGTTGAAAAATTTAATCAGGATCAAGAAATGAATAATAAAAA 2323  
Db |||||  
2708 TCCTCTTTCAAAATGAACCTTCTAAACTATTTGAATAAATAATATTTTGTGGAGAAAAA 2767  
Qy |||||  
2324 AGATGCA 2330  
Db |||||  
2768 AATTGAA 2774

## RESULT 22

AF332896  
LOCUS  
DEFINITION Streptococcus agalactiae strain M781 capsular polysaccharide biosynthesis gene cluster, partial sequence.  
ACCESSION AF332896  
VERSION AF332896.1 GI:18033333  
KEYWORDS  
SOURCE Streptococcus agalactiae  
ORGANISM Streptococcus agalactiae  
REFERENCE 1 (bases 1 to 4768)  
AUTHORS Kong, P., Gowan, S., Martin, D., James, G. and Gilbert, G.L.  
TITLE Serotype identification of group B streptococci by PCR and sequencing  
JOURNAL J. Clin. Microbiol. 40 (1), 216-226 (2002)  
MEDLINE 21635022  
PUBMED 11773119  
REFERENCE 2 (bases 1 to 4768)

AUTHORS Kong, P. and Gilbert, G.L.  
TITLE Direct Submission  
JOURNAL Submitted (02-JAN-2001) CIDMLS, ICPMR, Westmead Hospital, Darcy Rd., Sydney, NSW 2145, Australia  
FEATURES Location/Qualifiers  
source 1. 4768  
/organism="Streptococcus agalactiae"  
/mol\_type="genomic DNA"  
/strain="M781"  
/serotype="III"  
/db\_xref="taxon:1311"  
<1..126  
/gene="cpsD"  
1. 126  
/gene="cpsD"  
1. 126  
/gene="cpsD"  
/codon\_start=1  
/transl\_table=11  
/product="CpsD"  
/protein\_id="AAL57071.1"  
/db\_xref="GI:18033334"  
/translation="AKEQMEQSGKFLGIILNKVNESVATYGYNGYKRDKRK"  
178..1527  
/gene="cpsE"  
178..1527  
/gene="cpsE"  
/codon\_start=1  
/transl\_table=11  
/product="CpsE"  
/protein\_id="AAL57072.1"  
/db\_xref="GI:18033335"  
/translation="MIQTVVYFASLTLTITNFKSNKDLLFVLLIHYIVYLSDF  
YRDFWSGYLEEFKMKVYFYIPISSLFPIFNKSFTRLSFTFIANNLILYL  
LNSFLKYRYKSDTKVLTITNKSLSKTRFKYDHYIACVILDSSEKOCY  
DKNSLR1INKDALISLTCLTVDOAFINPIELFGYQIQDIINDIEMGVIVN  
VEALSFNIGKRIQTEGYSVITYSMKFYKSLAKRFLDITGAILIGLIGV  
FLVQIRKGGPAPFQNSVGRNRIFFPYKFSRMRVDAEQIKDLLVHNGMTLMFK  
LEDDPRITKIKRISIDELPQFYNVLKGDMSLVGRPTVDEYKYNSTQKRRLS  
FKPGITGLWQISGRNITDFDEIVKLQVYINWSIDIKIILLTLKVLVLTGAK"  
1551..2000  
/gene="cpsF"  
1551..2000  
/gene="cpsF"  
/codon\_start=1  
/transl\_table=11  
/product="CpsF"  
/protein\_id="AAL57073.1"  
/db\_xref="GI:18033336"  
/translation="MKICLVGSSGHLAHLNLLKPIWEKEDRFWFTDKEDARSILRE  
EIVYHCFPTNRNKNLVKNLTILAPKVLKERPDVYISSGAAVAPFPYIGKJGCKT  
VYISVFDRIKPTLTGKLVPVPTDKFIVQWEEKVYPKAINLGIF"  
2000..2473  
/gene="cpsG"  
2000..2473  
/gene="cpsG"  
/codon\_start=1  
/transl\_table=11  
/product="CpsG"  
/protein\_id="AAL57074.1"  
/db\_xref="GI:18033337"  
/translation="MIFVTGTHEQQFNRLIKVDRLLKGTGAIDQEVFIQYSDFFP  
QNCQSKFLSYDDMNSVMKEAEIVITHGGPATFMSVLSGLKLPVVPVRKQFGEHND  
HQIQPLKKIAHLYPLAWIEDVDVGLAELKRNIAIEKYQGNNDMFCHEKLEKIIGI"  
2470..3615  
/gene="cpsH"  
2470..3615  
/gene="cpsH"  
/codon\_start=1  
/transl\_table=11  
/product="CpsH"  
/protein\_id="AAL57075.1"  
/db\_xref="GI:18033338"  
/translation="MRKYLDLDSLFVALWVILVPOWQYOFLLIITIVLLLWKSEF  
RISISNSILFLWFIYLFALLIRGTQEDITFQRTIAELLKLISTGVALFFYNRYK

ADPNSVVRNVKVNVLFLITLVLYLFFPMLKPTLFCRELFSIEWPFFHMRIRLAAYF  
EYATLQGFILPSYPIFLPKQKHMENILISLLITICSYSGARILLVCMVLASLL  
LDYTLFTNLKLTKKNTFILGCTFELTACFSYNSIIIEKI IWRNQSTTRIMVYQ  
ESIIEVLKGNLFCQGIRIPSGEFLGSHSTYISVFVTSLLGVLVFSFALLIYKE  
AISKNYKIRLFFYTLCCYTLPEIDPNHNSIVLLSFTEFGIVGRKK"  
3612..4580  
/gene="cpsI"  
3612..4580  
/gene="cpsI"  
/codon\_start=1  
/transl\_table=11  
/product="CpsI"  
/protein\_id="AAL57076.1"  
/db\_xref="GI:18033339"  
/translation="MKRKVTVIPIYNSAYLKECVQSVLQOQTHPLIEVILLIDGSDT  
MSGEICDNLQEDNRILVFKHKGVSARNLGLDKSGEFTFVSDDFVAPNNIEI  
MLKLNITENADIAEVDPIISNERDYRKRRNFYKFKNNNSLKEFLSGNVEIVCT  
KLYKSIIGNLRFDENLKI GEDLLPNCCKLLCQEHRIVVDVTSSTYTRIVKTSVNNQK  
FENSLDFITIFNEISSIPARLANYVEAKFLREKI KCLRWFELSGNSIDNKIKVORE  
IFFKDIKSYFPYKAVKYLKLLSFLYLMKCSPLKYVMAYRRFQK"  
4614..>4768  
/gene="cpsJ"  
4614..>4768  
/gene="cpsJ"  
/codon\_start=1  
/transl\_table=11  
/product="CpsJ"  
/protein\_id="AAL57077.1"  
/db\_xref="GI:18033340"  
/translation="MIKKIEKDLISVPIYNVEDYLVCEIESLIVQYRNIIEILLIN  
DGSTDNCA"

Query Match 10.4%; Score 726.8; DB 1; Length 4768;  
Best Local Similarity 59.2%; Pred. No. 5.1e-73;  
Matches 1366; Conservative 0; Mismatches 902; Indels 38; Gaps 6;  
Qy 16 TGGCATTATTTGATGATAGCAGTTGCAATTTCTGCAATCTTAACAAGTCATATACCAA 75  
Db 173 TAGCGATGATTCACAAAGTTGTAGTTATTTCTGCAAGTTTGACATTAACATTAATTA 232  
Qy 76 ATGCTGATTTAAATCG---TTCCTGGAATTTTATCATTAATGATGTTTCAATTTTGGAT 132  
Db 233 CTCCCAATTTTAAAGCAATAAAGATTTATTTGTTGTTCTATGATACATTAATTTGTTT 292  
Qy 133 TTTTATATCTCGTATGCCAGTTGCAATTTGATAGTATAGAGTAATCTGATAGATTTGAAA 192  
Db 293 TTTATCTTCTGATTTTACAGACACTTTTGGAGTCGTGGCTATCTTGAAGAGITTAATA 352  
Qy 193 AAACATTTAACTAGTATATATTTGCAATTTTCTTACGCGCAGTATCATTTTGTGTTGG 252  
Db 353 TGGTATTGAATACAGCTTTTACTATATTTTATCATCAAGTTTCAATTTTATTTTATTTA 412  
Qy 253 AGAATAATTCGCACTTTCAAGACGTGCGGTGATTTTACATTAATAAATCTTCGTTT 312  
Db 413 AAAACTCTTTTACAAACGACACGACTTTCCTTTTATCTTTTATGCTATGAATTCGATTT 472  
Qy 313 TGGTATACCTATTTAACTGATTTTAAAGCAGTTTAAAGTAGCTTCTTATTTTTCGACAA 372  
Db 473 TATTATATCTATTTGAATTCATTTTAAATAATTTATCGAAAAATATTTCTTACGCTAAGTTT 532  
Qy 373 TCTATCAAAAAAAGACGATTTCTAATTTACACGGCTGACGATGGGAAATATGCAAGTTT 432  
Db 533 CACGAGATACCAAGTTGTTTGTATACGATAGGATTTCTTTCAAAATGACCTTTA 592  
Qy 433 TATTGTAATCATATAACAAATTCAAAAAATCTTGTGCAATGGTAGTATTTTAGTACAG 492  
Db 593 GGAATAAAT---ACGACCATTAATATATCGCTGCTGATCTTGTGACTCTCTGAAAAAG 649  
Qy 493 AAATAGATAAATAATTTATCATTTACCGCTCTATTATTCTGTGGAAGACGTATAGAGT 552  
Db 650 ATTGTTATGATTTGAAACATAAATCTGTTAAGGATAATAAACAAGATCTCTTACTTACG 709

ORIGIN  
Query Match 10.4%; Score 726.8; DB 1; Length 4768;  
Best Local Similarity 59.2%; Pred. No. 5.1e-73;  
Matches 1366; Conservative 0; Mismatches 902; Indels 38; Gaps 6;  
Qy 16 TGGCATTATTTGATGATAGCAGTTGCAATTTCTGCAATCTTAACAAGTCATATACCAA 75  
Db 173 TAGCGATGATTCACAAAGTTGTAGTTATTTCTGCAAGTTTGACATTAACATTAATTA 232  
Qy 76 ATGCTGATTTAAATCG---TTCCTGGAATTTTATCATTAATGATGTTTCAATTTTGGAT 132  
Db 233 CTCCCAATTTTAAAGCAATAAAGATTTATTTGTTGTTCTATGATACATTAATTTGTTT 292  
Qy 133 TTTTATATCTCGTATGCCAGTTGCAATTTGATAGTATAGAGTAATCTGATAGATTTGAAA 192  
Db 293 TTTATCTTCTGATTTTACAGACACTTTTGGAGTCGTGGCTATCTTGAAGAGITTAATA 352  
Qy 193 AAACATTTAACTAGTATATATTTGCAATTTTCTTACGCGCAGTATCATTTTGTGTTGG 252  
Db 353 TGGTATTGAATACAGCTTTTACTATATTTTATCATCAAGTTTCAATTTTATTTTATTTA 412  
Qy 253 AGAATAATTCGCACTTTCAAGACGTGCGGTGATTTTACATTAATAAATCTTCGTTT 312  
Db 413 AAAACTCTTTTACAAACGACACGACTTTCCTTTTATCTTTTATGCTATGAATTCGATTT 472  
Qy 313 TGGTATACCTATTTAACTGATTTTAAAGCAGTTTAAAGTAGCTTCTTATTTTTCGACAA 372  
Db 473 TATTATATCTATTTGAATTCATTTTAAATAATTTATCGAAAAATATTTCTTACGCTAAGTTT 532  
Qy 373 TCTATCAAAAAAAGACGATTTCTAATTTACACGGCTGACGATGGGAAATATGCAAGTTT 432  
Db 533 CACGAGATACCAAGTTGTTTGTATACGATAGGATTTCTTTCAAAATGACCTTTA 592  
Qy 433 TATTGTAATCATATAACAAATTCAAAAAATCTTGTGCAATGGTAGTATTTTAGTACAG 492  
Db 593 GGAATAAAT---ACGACCATTAATATATCGCTGCTGATCTTGTGACTCTCTGAAAAAG 649  
Qy 493 AAATAGATAAATAATTTATCATTTACCGCTCTATTATTCTGTGGAAGACGTATAGAGT 552  
Db 650 ATTGTTATGATTTGAAACATAAATCTGTTAAGGATAATAAACAAGATCTCTTACTTACG 709





source	1. .4768 /organism="Streptococcus agalactiae" /mol_type="genomic DNA" /strain="SG99/143" /serotype="III" /db_xref="taxon:1311" <1. .126 /gene="cpsD" /gene="cpsD" /codon_start=1 /transl_table=11 /product="CpsD" /protein_id="AAL56360.1" /db_xref="GI:18029140" /translation="AKEQMEQSGSKFLGIILNKVNESVATYGDYGNVGRDRKKR" 178. .1527 /gene="cpsE" 178. .1527 /gene="cpsE" /codon_start=1 /transl_table=11 /product="CpsE" /protein_id="AAL56361.1" /db_xref="GI:18029141" /translation="MIQTVVYVFSASLTLTITPNFKSNKDLLFVLLIHIVFYLSDF YRDFSRGLREFKMWLVKSYFYIIFISSLLFIPKNSFTTRLSFFFIAMNSILLVL LNSFLKYRKYSYAKESRDTKVLITNKDSLKMTFRNKYDHNVIANCILDSSEKCY DLKNSLIRNKDALSELCTLVQDQFINPIELFKYQIQDINDIANGVIVNVN VEALSFNIGSKRIQTGEGSYITYSMKFKYSHLIARFLDITGAILGLICGIVAI FLVQIRQDGKGPALFQSNVRNGRIIFRFYKPSMRVDAEQIKKDLVHNGMTGLMFK LEDDPRITKIGKFKRTSIDELPOFYNVLKGDSILVGRPTVDYKYNSTOKRRLS FKPGITGLWQISGRNNITDFDEIVKLDVQYINEMSIMSDIKIILLTLKVVLLGTGAK" 1551. .2000 /gene="cpsF" 1551. .2000 /gene="cpsF" /codon_start=1 /transl_table=11 /product="CpsF" /protein_id="AAL56362.1" /db_xref="GI:18029142" /translation="WKICLVGSSGSHLAHLNLLKPIWEKEDRFWTFDKEDARSILRE BIVVHCRPTNRNVKNLKNITILAFKVLKRPDVIISGGAUAVPPFYIGKLGCKT VYIEVPRIDKPTLTGLKLVVPYTDKFIQWHEMKVYPKAINLGGIP" 2000. .2473 /gene="cpsG" 2000. .2473 /gene="cpsG" /codon_start=1 /transl_table=11 /product="CpsG" /protein_id="AAL56363.1" /db_xref="GI:18029143" /translation="MIFVTVGTHEQQFNRLIKEVDRLKGTGTAIDOEVIQIGSDPEP QNCWSKELSDVNMVMEAEIVITHGCPATMSVLSGLKLVVPRKOPGEGHND HQIQPKKIAHLPLAWIEDVDGLAALKRNIAIEKYQGNNDMFCHKLEKLEIGI" 2470. .3615 /gene="cpsH" 2470. .3615 /gene="cpsH" /codon_start=1 /transl_table=11 /product="CpsH" /protein_id="AAL56364.1" /db_xref="GI:18029144" /translation="MRKYLDDYSLFVALWLVLPVNMVQFLIITIIVLLLMKSEF RISNSNLSILFLMLFYLPAILINGTQEDITFPQPIAELLKLISGTGVALPYNRYK ADFNSSVRNVVNVVFLVILTYLFFMLKPTLFGRELFSIEWFPHMRIRLAAYF EYATILGQFLFSYLPILFKPQKHMENILISLLTFCISYFGSARLLVCMVLVASLL LDYILFKNLKILKMTFLGTFITACFSYNIWSIIKIIIMRNGSTITRMIVYQ ESIIEVLGNILFGQIRIPSEGIIFLGSHSTYISVFYRTSLIGLIVLFSAPILLYKE AISKNYKIYRFPFYLLCYLPFEEIDPNHWSIVLLFSFPGIVGRACK"
gene	3612. .4580 /gene="cpsI" 3612. .4580 /gene="cpsI" /codon_start=1 /transl_table=11 /product="CpsI" /protein_id="AAL56365.1" /db_xref="GI:18029145" /translation="MKEKVTVIIPIYNSEAYLKECVOSVLOOHTPLIEVILIDGSDT NSGEICDNLQEDNRILVPHKKGVSARNIGLDKSTGEFITFVDSDFVAPNMIEI MLNLTENADIAREVDPIISNERDYRKKRPNFYKPNNSLKEFLSGNVAENIVCT KLKYSIIGNLPDENLIGEDLLFNCKLJCQEHRIVVDITSSLTITRYKIVSNQK IFNFDLFIITFNEISLAPLANLYAEKFLREKIKLRKMFELGNSIDNKIKVQRE IFFKDIKSYFPYKAVKYLKSLKLLSFYLMKSPKLYMAYRRFQK"
gene	4614. .>4768 /gene="cpsJ" 4614. .>4768 /gene="cpsJ" /codon_start=1 /transl_table=11 /product="CpsJ" /protein_id="AAL56366.1" /db_xref="GI:18029146" /translation="MIKKIEKDLISVIVPIYNVEDYLVECTIESLIVQYRNIIEILLIN DGSNDNCA"
CDS	Query Match 10.4%; Score 726.8; DB 1; Length 4768; Best Local Similarity 59.2%; Pred. No. 5.1e-73; Matches 1366; Conservative 0; Mismatches 902; Indels 38; Gaps 6; QY 16 TGGCATTATTGATGATAGTAGCAGTTGCAATTTCTGCAATCTTAACAAGTCATATACCAA 75 DB 173 TAGCGATGATTCAAACAGTTGTAGTTATTTTCTGCAAGTTTGACATTACATTAA 232 QY 76 ATGCTGATTTAAATCG---TTCTGGAATTTTATCATAAATGATGGTTCATTATTTTGCAT 132 DB 233 CTCCCAATTTTAAAGCAATAAAGATTTATTTGTTGTTCTATTGATACATTATATTTT 292 QY 133 TTTTATATCTCGTATGCCAGTTGAAATTTAGTATAGAGTAAATCTGATAGAGTTTGA 192 DB 293 TTTATCTTCTCGATTTTACAGACACTTTTGGAGTCGTGGGTATCTTGAAGAGTTTAA 352 QY 193 AAACATTAACTAGTATATATTTGCAATTTTCTTACGCGCAGTATCATATTTTGTGG 252 DB 353 TGGTATTGAAATACAGCTTTTACTATATTTTCAATCAAGTTTCATATTTTATTTT 412 QY 253 AGAATAATTTCCGCACTTTTCAAGACGTGGTCCGCTGATTTTCCACATTAATAAATCTCGTTT 312 DB 413 AAAACTCTTTTACAACGACACGACTTTTCTCTTTTACITTTTATTCCTATGAATTCGATTT 472 QY 313 TGGTATACCTATTTAACGTAATTAATTAAGCAGTTTAAGGATAGCTTTCTATTTTTCGACA 372 DB 473 TATTATATCTATGAAATTCATTTTAAATATATTATCGAAAAATATTTCTTACGCTAAGTTT 532 QY 373 TCTATCAAAAAGACAGATCTTAATTACAACGCTGAACGATGGGAAATATGCAAGTTT 432 DB 533 CACGAGATACCAAGTTGTTTGTGATACGAATAAGGATCTTTATCAAAAATGACCTTTA 592 QY 433 TATTGAAATACATAAAACAAATTTCAAAAAATCTTGTGTCATTGGTAGTTTGTAGGTACAG 492 DB 593 GGAATAAT---ACGACCATTAATTATATCGCTGCTGTATCTTGGACTCTCTGAAAAGG 649 QY 493 AAATAGATAAAATAATTTATCATTTACCGCTCTATTATTCTGTGTGGAAGACGTATAGAGT 552 DB 650 ATTGTTATGATTTGAAACATAACTCGTTAAGGATAATAAACAAGATGCTCTTACTTCAG 709 QY 553 TTTCAACAGGGAAGTGTGCGACACGCTCTTTATATAAATCTACCAAGTGNATTTTAGAGC 612 DB 710 AGTTAACCTGCTTAACCTGTCAGCTTTTATTAACACCCCATCCCAATTGAATTTTGGTA 769 QY 613 TAAAGCAA-----TTTCGTTTTCAGATTTTGTAGTTGTTGTTAGTATTGATGTAAGCGTTG 663

Db 770 AATACCAATACAAAGATATATTAATGACATGGAAGCAATGGGAGTGAATGTCATGTTA 829  
Qy 664 ATATTAATTCANTCGGTTTACTCGGTGAAAAACAAAAATCCAACTGCTAGGTGACC 723  
Db 830 ATGTAGAGGCACTTAGCTTTGATATATAGGAGAAAGCGAATCCAAACTTTTGAAGGAT 889  
Qy 724 ATAGCATTTGAACCTTTTCCAAATTTTATAGAGCTGATCATATCATGATGAAGAC 783  
Db 890 ATAGTGTTATACATATCTATGAATTTCTATAAATATAGTCACCTTATAGCAAAACGAT 949  
Qy 784 TTTTGGATATACCTCGAGCGGTAGTCGGGTAAATTTTGTGGTATAGTTTCTATTTTGT 843  
Db 950 TTTTGGATATACCGGTGCTATATAGTTTGTCTCATATGGCATTTGTGGCAATTTTC 1009  
Qy 844 TAGTTCAAATTTATCGTAGAGATCGGTGACCGGCTATTTTGTGTCAGAAAACGAGTTGGAC 903  
Db 1010 TAGTTCCGCAATCAGAAAAGATGGTGAACGGGCTATCTTTCTCAAAATAGAGTAGGTC 1069  
Qy 904 AGAATGACGCATATTTACATTTCTACAAGTTTCAGTTCGATCGATGTTGTTGATGCTGAGGAC 963  
Db 1070 GTAATGGTAGGATTTTATAGATTTCTATAATTCAGATCAATGCGAGTAGATGCGAAACAA 1129  
Qy 964 GCAAAAAGACTTCTCAGCCAAAACAGATGCAAGGTGGGTATGTTTAAAAATGGAA 1023  
Db 1130 TTAGAAGATTTATAGTTTCAATCAATCAATGACAGG-----CTAATGTTTAAAGTTAGA 1184  
Qy 1024 AAACGATCCTAGAAATTTACTCCAAATGGAATTTTCATAGCAAAAACAAAGTTTACAGAGT 1083  
Db 1185 AGATGATCCTAGAAATTTACTAAAATAGGAAAATTTATTTCG-AAAAACAAGCATAGATGAGT 1243  
Qy 1084 TACCACAGTTTATAATGTTTAAATGCGGATATAGTCTAGTGTGTTGATCAGGTCACCTA 1143  
Db 1244 TGCCTCAATTTCTAATGTTTAAAGGTGATATGRTTTTAGTAGGAACACGCGCTCCCA 1303  
Qy 1144 CAGTTGATGAATTTGAAAAATATATCTCTGCTCAAAAGAGACGATTCAGTTTAAACAG 1203  
Db 1304 CAGTTGATGATATGAAGAATTAATTTCAACGAGAGAGCGCTTAGTTTAAAGCAG 1363  
Qy 1204 GGATTACAGGTCCTGCGAGGTAGTGTGCTAGTATATATCAACAGACTTCGACGACGTAG 1263  
Db 1364 GAATCACTGGTTTGTGCAAAATATCTGTAAGAAATATATTAATCTGATTTTGAATAATCG 1423  
Qy 1264 TTCGGTTGGACTTAGCATACATTAATTTGGACTATCTGGTCAGATTAATAATTTTAT 1323  
Db 1424 TAAAGTTAGATGTTTCAATATATCAATGAATGGTCTATTGGTTCAGATTAAGAATTATTC 1483  
Qy 1324 TAAAGACAGTGAAAGTTGTTATGTTGAGAGAGGGAAGTAAGTAAAGTA----- 1372  
Db 1484 TCCTAACTAAAGTAGTTTACTCGGACAGAGCTAGTAAGGTAAGGTTTGAAG 1543  
Qy 1373 -----TATGAAGTTTGTGTTGTCGTTTCTCAAGGGGACATTTGACTCACTGTATTT 1426  
Db 1544 GAATATAATGAAATTTGCTGTTGTTCAAGTGGTGGTTCATCTAGCACACTTGAACCT 1603  
Qy 1427 GTTAAACCGTTTGGAGGAGAGAACGTTTTTGGGTAAACATTTTCATAAGAGGATGC 1486  
Db 1604 TTTGAAACCCATTTGGGAAAAGAGATAGGTTTTTGGGTAACTTTGATAAAGAGATGC 1663  
Qy 1487 AAGAAGTCTTTTGAAGATGAAAAATGATCTCCATGTTTACTTCCAAACAAATGCAATCT 1546  
Db 1664 TAGAGTATTTCTAGAGAGAGATGTTATATCATTTCTTTCCAAACAAACCGTAATGT 1723  
Qy 1547 CATTAATTTAGTGAATAATCTTTCTTAGCTTTTCAAAATTTTACGTGATGAGAAACAGA 1606  
Db 1724 CAAAAACTTGTGTAATAATACTATTCTAGCTTTTAAAGTCTCTAGAAAAGAAAGACAGA 1783  
Qy 1607 TGTATTATTTCATCTGTTGGCGGTTGCTGTCCTCTTTTACATCGGAAACHTATT 1666  
Db 1784 TGTATTATATATCATCTGTTGGCGCTGTAGCAGTACCATTTCTTTTATATGTTAGTATTAT 1843  
Qy 1667 TGGAGCAAGACGATTTATATTTGAAGTATTGATCGAGTTTAATAAATCTACATTAACCTGG 1726

Db 1844 TGGTGTTAAGACCGTTTATATATAGAGGTTTTTCGACAGATAGATAAACCAACTTTGCACAG 1903  
Qy 1727 AAACTAGTTTATCCGTTAAACAGATATTTTATTTGTTTTCAGTGGGAGAAATGAAAGGT 1786  
Db 1904 AAAATTAGTGATCTCTGTTAAACAGATAAATTTATTTGTTTCAGTGGGAGAAATGAAAAAGT 1963  
Qy 1787 ATATCTTAATCTATTAACTTTGGGAGTATTTTAAATGATTTTGTAAACAGTAGGAAC 1846  
Db 1964 TTATCTTAAGGCAATTAATTTAGGAGAAATTTTAAATGATTTTGTTCAGCTGGGAGCA 2023  
Qy 1847 CATGAACAAACAGTTTAAATCGATTGATAAAGAGATTTGATTTATGAAAAAATGGAAGT 1906  
Db 2024 CATGAACAGCAGTTCAACCGTCTTTATTAAGAGAGTTGATAGATTAAAAGGACAGAGTGCT 2083  
Qy 1907 ATAACCGACGAATATTTATTTCAACAGAGATTTCTGACTATATTTCCAGATATTTGCAAG 1966  
Db 2084 ATTGATCAAGAGTGTTCATTTCAACGGGTTTACTCAGACTTCGAACCTCAGAAATTTGTGAG 2143  
Qy 1967 TATAAAAAATTTCTCAGTTTCAAGAAATGGAACAATATATTAACRAATCAGAGTAGTT 2026  
Db 2144 TGGTCAAAATTTCTCTCATATGATGATGAATCTCTTACATGAAGAAAGCTGAGATTGTT 2203  
Qy 2027 ATTGTCACGAGGCGCCGCTACTTTTATGAATTCATTTATCCAAAGGAAAAACAATTA 2086  
Db 2204 ATCACACATGGCGCCGACGCGTTTATGTGTCAGTTATTTCTTTAGGNAATTTACCAGTT 2263  
Qy 2087 TTGTTTCTTAGACAAAAAAGTATGTTGAACATGTAAATGATCATCAAGTAGAGTTGTA 2146  
Db 2264 GTTGTTCCTAGGAGAAAGCAGTTTGGTGAACATATCAATGATCATCAAAATCAAAATTTTGA 2323  
Qy 2147 AGAAGAAATTTTACAAGTAAATATATTTTATTTATAGAAATATAGATGATTTGTTGAA 2206  
Db 2324 AAAAAATTTGCCACCTGATCCCTGGCTTGGATTGGAATGATGAGACTTTGCGGAA 2383  
Qy 2207 AAAATTTTGAAGTTTCTAAGCAAACTAATCTTACATCAATTAATATTTTGTGAA 2266  
Db 2384 GCCTTGAAGGAATATAGCTACAGAAAATATCAGGAAATATGATATGTTTGTGTCAT 2443  
Qy 2267 AGATTAACAAATAGTTGAAAAAT 2292  
Db 2444 AAATTAGAAAAATTTAGTGAAAT 2469

RESULT 24  
AF363056 4768 bp DNA linear BCT 07-JAN-2002  
LOCUS Streptococcus agalactiae strain GB00/031 capsular polysaccharide  
DEFINITION biosynthesis gene cluster, partial sequence.  
ACCESSION AF363056  
VERSION AF363056.1 GI:18029147  
KEYWORDS Streptococcus agalactiae  
SOURCE Streptococcus agalactiae  
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.  
REFERENCE 1 (bases 1 to 4768)  
AUTHORS Kong,F., Gowan,S., Martin,D., James,G. and Gilbert,G.L.  
TITLE Serotype identification of group B streptococci by PCR and sequencing  
JOURNAL J. Clin. Microbiol. 40 (1), 216-226 (2002)  
MEDLINE 21635022  
PUBMED 11773119  
REFERENCE 2 (bases 1 to 4768)  
AUTHORS Kong,F. and Gilbert,G.L.  
TITLE Direct Submission  
JOURNAL Submitted (22-MAR-2001) CIDMLS, ICPMR, Westmead Hospital, Darcy Road, Sydney, NSW 2145, Australia  
FEATURES  
source Location/Qualifiers  
1. 4768  
/organism="Streptococcus agalactiae"  
/mol\_type="genomic DNA"  
/strain="GB00/031"  
/serotype="III"

```

/db_xref="taxon:1311"
<1..126
/genes="cpsd"
<1..126
/genes="cpsd"
/codon_start=1
/transl_table=11
/product="CpsD"
/protein_id="AAL56371.1"
/db_xref="GI:18029148"
/translation="AKEQEQSGSKFLGIILNKNVSVATYGDYGNVGRDRKRK"
178..1527
/genes="cpsE"
178..1527
/genes="cpsE"
/codon_start=1
/transl_table=11
/product="CpsE"
/protein_id="AAL56371.1"
/db_xref="GI:18029149"
/translation="MIQTVVVYFSAISLTLITPMPKSNKDLFLVLIHIVYVLSDF
YRDFSRGVEEFKMLVLYFYIYISSLFFIFKNSFTYFRLSFTFIAMNSILLYL
LNSFLYKRYKYAKFSDTKVVLITKDSLSKMTFRNKVDHNYIAVCLDSEKDCY
LNSFLIRLTKNDLSTELCTLDQAFINIPLEKGYQIOIDINDIEMGVNVN
VEALSPDNIGEKRIQTFEGSYVITYSMKFYKSHLIAGKPLDITGAILGICGIVAI
FLVPOIKDGGPAIFSQNRVGRNGRIFRYPKFRSMRVDAEQIKDILLVHQNMTGLMFK
LEDDPRITKIGIRIKTSIDELPQFYNLGDMSLVGTPTVDEYKYNSTQKRLS
FKPGITGLWQISGRNNITDFDEIVKLVDVQYINNEWSIWSIDKIILLTLKVLVLLGTAK"
1551..2000
/genes="cpsF"
1551..2000
/genes="cpsF"
/codon_start=1
/transl_table=11
/product="CpsF"
/protein_id="AAL56371.1"
/db_xref="GI:18029150"
/translation="MKICLVGSSGHLAHLNLLKPIWEKEDRFWTFDKEDARSILRE
EIVHCFPTNRRNVKNVNTILAFKVLKERPDVLISSGAAVAVPFFYIGKLFCKT
VYIEVDRIDKFLTKGLVYPTVDKRFVQWEMKQVYKPAINLGGIF"
2000..2473
/genes="cpsG"
2000..2473
/genes="cpsG"
/codon_start=1
/transl_table=11
/product="CpsG"
/protein_id="AAL56371.1"
/db_xref="GI:18029151"
/translation="MIFVTVGHEQPNRLIKEVDRLKGTGAIDDEVFTQYGSDFEP
QNCWKSFLSYDDMSYMKAEIIVITHGPAFMVSISLGLKPLVVPVPRKQFGEHND
HQIQPKIAHLVPLAWIEDVDGLAELKRNATEKYQGNNDMFCKLEKIIGEI"
2470..3615
/genes="cpsH"
2470..3615
/genes="cpsH"
/codon_start=1
/transl_table=11
/product="CpsH"
/protein_id="AAL56371.1"
/db_xref="GI:18029152"
/translation="WRKYLDDLSYLFYALWLVILVFNQWYFLIITIIIVLLMLKSEF
RISIGNSILFLWLFYILFALLIRGTQSDITFORFIELLKLIGTYALPFFVNYRK
ADFNSSVRNVVNVVFLITLVLYLPMLKPTLFGRELSIEWFPMRLRLAAYF
EYATLIGQFLSPYFLEKPKQKMHENILISLLITCSYFSGARILVCMVILLASFL
LDYILFKNLTKQNTFLGMTFLFIFACFSYNIWSILIEKLIIMRNQSTITRMIVYQ
ESIIVLKNLIFGGQIRIPSEGEIPLVGSHTYISVYRTSLGLVLYFSAPILLYKE
AISKNNKIYRLFYFYLICVTLFEEIDPNHWSIVLLFSTFGIVGRAK"
3612..4580
/genes="cpsI"
3612..4580
/genes="cpsI"
/codon_start=1

```

830 ATGTAGAGCGACTTAGCTTTTGATATATAGAGAAAGCGAATCCAACTCTTTGAAGGAT 889  
840 ATAGCATGTAACTTTTCCAAATTTTATTAAGCCCTAGTCTATATCATGATGAAGACGAC 783  
850 ATAGTGTATTACATATCTATGAAATCTATATAATATATAGTCACTTTATAGCAAAACGAT 949  
860 TTTTGGATATACCGGAGCGGTAGTCTGGGTTAAATTTCTGGGTATATAGTCTTCTATTTTGT 843  
870 TTTTGGATATACCGGAGCGGTATATATAGTCTTATATAGTCTTCTATATAGTCTTCTATTTTC 1009  
880 TAGTTCCTCAATTTATTCGTAGAGATCGTGGACCGGCTATTTTGTCTCAGAAAACGAGTTCGAC 903  
890 TAGTTCGCAAAATCAGAAAAGATGTGGACCGGCTATCTTTCTCAAAATATAGATAGTCTC 1069  
900 AGAATGAGCGCATATTTACATTTTACAAATTTTCAAGTTTTCGATGTATGTGTGCTGAGGAGC 963  
910 GTATATGCTAGGATTTTATAGTCTTATATAATTTTCAATTTTCAATTTTCAATTTTCAATTTT 1129  
920 GCAAAAAGACTTCTCAGCCAAACAGATGCAAGGCTGGGTATGTTTAAATGGGAA 1023  
930 TTAAGAAAGATTTATATAGTCTCACAATCAAAATGACAGGG-----CTAATGTTTAAATTTAGA 1184  
940 AAACGATCTAGATTTACTTCAATTTGACATTTTATACGCAAAACAAAGTTTATAGCAGT 1083  
950 AGATGATCTAGATTTACTTAAATTAGGAAATTTATTCG-AAAAACAAGCATAGATGAGT 1243  
960 TACCACAGTTTATATGTTTAAATGCGGATATGATCTAGTGTGTTGTTACAGTCCACTA 1143  
970 TGCCTCAATTTCTAATATGTTTAAAGGTGATATGATGTTTATAGTAGGAACACGCGCTCCCA 1303  
980 CAGTTGATGATTTGAAAATATATCTCTGTTCAAAAGACGATTCAGTTTAAACAG 1203  
990 CAGTTGATGATATGAAAAGTATATTTCAACGCAAGACGACGCTTAGTTTAAAGCCAG 1363  
1000 GGATTACAGGCTCTCGCAGGTTAGTGTCTGATGATATATACAGACATTCGACGACGTAG 1263  
1010 GNATCATGTTTGTGGCAATATCTGTAGAAATATATTTACTGATTTTGTGAAATCG 1423  
1020 TTGCGTTGGACTTAGCATACATTCGATTAATTTGGACTATCTGTTGATATATTAATTTTAT 1323  
1030 TAAAGTTAGATGTTTCAATATATCAATGAATGTTCTATTTGTTGATATTAAGATTTATC 1483  
1040 TAAGACAGTCAAGTTGTTATTTGTTAGAGAGGAGTAAAGTAAAGTAAAGTAAAGTAAAG 1372  
1050 TCCTAAACATTAAGGTTAGTTTACTCGGAGCAGGACGATTAAGTAAAGTAAAGTAAAG 1543  
1060 -----TATGAAAGTTTGTGTTGTTGTTCTTTCAGGCGGACATTTGACTCACTTGTATTT 1426  
1070 GNATATATGAAATTTGTTGTTGTTGTTCAAGTGGTGGTCACTAGCACACTTGAACCT 1603  
1080 GTTAAACCGTTTGGAGGAGGAAAGACGTTTGTGGTTAACTTTCGATTAAGAGGATGC 1486  
1090 TTTGAAACCGATTTGGGAAAGGAAAGATAGTTTGTGGTAACTTTCGATTAAGAGGATGC 1663  
1100 AAGAGTCTTTGAGATGAAAGAAATGATTCATGTTACTTCCACAAATCGCATCT 1546  
1110 TAGGATATTTCTAAGAGAGAGATTTGATATCATTTGCTTCTTCCAAACAAACCGTAAATGT 1723  
1120 CATTAATTTAGTGAATAATCTTTCTTAGCTTTTCAAAATTTTACGTGATGAGAAACAGA 1606  
1130 CAATACTGTTGTAATAATACTATTTCTAGCTTTTAAAGTCTTGAAGAAAGAGACAGGA 1783  
1140 TGTATTTATTTCACTGTTGCGCGGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1666  
1150 TGTATCATATCATCTGTTGCGGCTGTAGCAGTACCATTCTTTTATATTTGTTAGTTTATT 1843  
1160 TGGAGCAAGACGATTTATTTAGATTTTATGATGCGAGTTTAAATTTCAATTTAACTGCG 1726  
1170 TGGTTGTAAGACCGTTTATATAGAGGTTTTCGACAGGATAGATTAACCACTTTTGCAGG 1903  
1180 AAAACTAGTTTATCCCGTAACAGATATTTTATTTGTTTCAAGTGGGAGAAATGAAAGAGGT 1786  
1190 AAAATTAGTGTATCTCTGTAACAGATTAATTTATTTGTTTCAAGTGGGAGAAATGAAAGAGT 1963

1787 ATATCTTAATCTATTAATCTTGGGAGTATTTTTTAAATGATTTTTTGTAAACAGTAGGAAT 1846  
1797 TTAATCTTAAGGCAATTAATTTAGGAGGAATTTTTTAAATGATTTTTTGTACAGATGGGACA 2023  
1807 CATGAACAACAGTTTAAATCGATTGATAAAGAGATTTGATTTATTTGAAAAAATTTGGAAGT 1906  
1817 CATGAACAACAGTTTAAATCGATTGATAAAGAGATTTGATTTATTTGAAAAAATTTGGAAGT 2083  
1827 ATAAACCGACGAAATATTTTAAACAGGATATTTCTGACTATATTTCCAGAAATTTGCAAG 1966  
1837 ATTGATCAAGAAGTGTTCATTTCAACCGGTTACTCGACTTTCGAACCTCGAATTTGTCAG 2143  
1847 TATAAAAAATTTCTAGTTTCAAAAGAAATGGAACAATATATTTAAACAATCAGAAGTAGTT 2026  
1857 TGGTCAAAATTTCTCTCATATGATATGAACTTTTACATGAAGAAGCTGAGATTTGTT 2203  
1867 ATTTGCGACGAGCGCCGCTACTTTTATGAAATTTTATTTTCAAAAGGAAAAACAATTA 2086  
1877 ATCACAATGCGCGCCGAGCGTTTATGCTAGTTTATTTCTTTAGGGAATTTACCAAGTT 2263  
1887 TTGTTTCTTAGACAACAAAAAGTATGTTGAACATGTAATGATCATCAAGTAGAGTTTGT 2146  
1897 GTTGTCTTAGGAGAAAGCAGTTTGTGTAACATATCAATGATCATCAATTAATTTT 2323  
1907 AGAAGAAATTTTACAAGATTAATATTTTATTTTATGAAAAATATAGATGATTTGTTGAA 2206  
1917 AAAAAAATTTGCCACCTGTATCTTGGCTTGGATTTGAAGATGATAGTGGACTTTCGGA 2383  
1927 AAAATTTATGAAAGTTTCTAAGCAAACTAACTTTTACATCAATTAATTTTGTGTTGAA 2266  
1937 GCCTTGAAGAGGAAATAGCTACAGAAAAATATCAGGAAATTAATGATATGTTTGTGAT 2443  
1947 AGATTTAAACAAATAGTTGAAAAAT 2292  
1957 AAATTAGAAAAATTTAGTGGAAT 2469

## RESULT 25

AF163833 LOCUS 17276 bp DNA linear BCT 03-AUG-2000  
DEFINITION Streptococcus agalactiae capsular polysaccharide synthesis locus, complete sequence.  
ACCESSION AF163833 L09116  
VERSION AF163833.1 GI:5823209  
KEYWORDS Streptococcus agalactiae  
SOURCE Streptococcus agalactiae  
ORGANISM Streptococcus agalactiae  
REFERENCE 1 (bases 4755 to 6119)  
AUTHORS Rubens, C.E., Heggen, L.M., Haft, R.F. and Wessels, M.R.  
TITLE Identification of cpsB, a gene essential for type III capsule expression in group B streptococci  
JOURNAL Mol. Microbiol. 8 (5), 843-855 (1993)  
MEDLINE 93360815  
PUBMED 8355611  
REMARK cpsB now called cpsE  
REFERENCE 2 (bases 15471 to 16670)  
AUTHORS Haft, R.F., Wessels, M.R., Mebane, M.F., Conaty, N. and Rubens, C.E.  
TITLE Characterization of cpsF and its product CMP-N-acetylneuraminic acid synthetase, a group B streptococcal enzyme that can function in K1 capsular polysaccharide biosynthesis in Escherichia coli  
JOURNAL Mol. Microbiol. 19 (3), 555-563 (1996)  
MEDLINE 96228704  
PUBMED 8830246  
REMARK cpsF now called neuA  
REFERENCE 3 (bases 1 to 17276)  
AUTHORS Chaffin, D.O., Beres, S.B., Yim, H.H. and Rubens, C.E.  
TITLE The serotype of type Ia and III group B streptococci is determined by the polymerase gene within the polycistronic capsule operon  
JOURNAL J. Bacteriol. 182 (16), 4466-4477 (2000)  
MEDLINE 20372630





QY	1787	ATATCCTAAATCTATTAACTTCGGGAGATATTTTTTAATGATTTTTTGTAAACAGTAGGAACT	1846
Db	6556	TTATCCTAAGGCAATTAATTTAGGAGGAATTTTTTAATGATTTTTTGTACAGTGGGACA	6615
QY	1847	CATGAACAACAGTTTAAATCGATTGATAAAGAGATTGATTTATCGAAAATAAATCGAAGT	1906
Db	6616	CATGAACAGCAGTTCAACCGTCTTATTAAGAGAGTTGATAGATTAAAGGGACAGTGCT	6675
QY	1907	ATAACCGACGAATATTTATTCAAAACAGGATATTTCTGACTATATTTCCAGAATATTTGCAAG	1966
Db	6676	ATTGATCAAGAAGTGTTCATTCAACCGGTTACTCAGACTTCGAACCTCAGAATTTGTCAG	6735
QY	1967	TATAAAAAATTTCTCAGTTACAAAGAAATGGACAAATATATTAAACAATCAGAAGTAGTT	2026
Db	6736	TGGTCAAAAATTTCTCTCATATGATGATATGAACTCTTTACATGAAAAGAGCTGAGATTGTT	6795
QY	2027	ATTTGCCACGAGGCCCGCTACTTTTATGAATTCATTATCCAAAGGAAAACAAATTA	2086
Db	6796	ATCACACATGGCGGCCCGACGTTTATGTCAGTTATTTCTTAGGGAAATTTACCGTT	6855
QY	2087	TTGTTTCTTAGACAAAAAAGTATGGTGAACATGTAATGATCATCAAGTAGAGTTTGTA	2146
Db	6856	GTGTTCTCTAGGAGAAAGCAGTTTGGTGAACATATCAATGATCATCAATACAATTTTA	6915
QY	2147	AGAAGAAATTTACAAGATAATAATATTTTATTATAGAAAATATAGATGATTTGTTGAA	2206
Db	6916	AAAAAAATTGCCCACTGTATCCCTTGGCTTGGATTGAAGATGTAGATGGACTTCGGAA	6975
QY	2207	AAAAATTATGAAGTTTCTAAGCAACTAATTTACATCAATATAATTTTTTTGTGAA	2266
Db	6976	GGTTGAAAAGGAATATAGCTACAGAAAATATCAGGGAATAATGATATGTTTGTGAT	7035
QY	2267	AGATTAAAAACAATAGTTGAAAAAT	2292
Db	7036	AAATTAGAAAAAATTTATAGTCAAT	7061

Search completed: April 29, 2005, 10:37:38  
Job time : 28940 secs



This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 28, 2005, 15:28:57 ; Search time 3184 Seconds  
(without alignments)  
12999.633 Million cell updates/sec

Title: US-09-767-041-9  
Perfect score: 6992  
Sequence: 1 atcgccaaacgaattggca.....gcaaacagcaaaagaagctt 6992

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870567 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*

- 1: Geneseq1980s:\*
- 2: Geneseq1990s:\*
- 3: Geneseq2000s:\*
- 4: Geneseq2001as:\*
- 5: Geneseq2001bs:\*
- 6: Geneseq2002as:\*
- 7: Geneseq2002bs:\*
- 8: Geneseq2003as:\*
- 9: Geneseq2003bs:\*
- 10: Geneseq2003cs:\*
- 11: Geneseq2003ds:\*
- 12: Geneseq2004as:\*
- 13: Geneseq2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6992	100.0	6992	3	Aaz60930 Nucleotid
2	1195.8	17.1	26281	3	Aaz60929 Nucleotid
3	995.6	13.7	17468	6	Aba01441 Streptoco
4	889	12.7	16032	13	Adsl3200 S. thermo
5	798.8	11.4	110000	6	Continuation (12 o
6	727.8	10.4	6865	6	ABK90550 Betal, 3-g
7	726.8	10.4	17276	8	ACA64723 Streptoco
8	726.8	10.4	17276	10	ADF43363 Streptoco
9	721.4	10.3	25020	12	ADO40235 S. agalac
10	649.2	9.3	2226	11	ADM79774 Group B S
11	428.4	6.1	6850	4	Aah43197 Portion o
12	423.2	6.1	1368	4	Aas55950 Streptoco
13	415.8	5.9	1443	13	ADR93668 Novel S.
14	383.2	5.5	18274	6	Aba01446 Streptoco
15	376.2	5.4	19966	3	Aaz30355 Nucleotid
16	376.2	5.4	19966	3	Aaz45258 DNA encod
17	376	5.4	28170	6	Aba01447 Streptoco
18	373	5.3	15962	6	Aba01445 Streptoco
19	371.6	5.3	20555	6	Aba01443 Streptoco
20	371.4	5.3	18488	6	Aba01444 Streptoco

21	371.4	5.3	19738	6	ABA01436	Aba01436 Streptoco
22	366	5.2	25243	6	ABA01442	Aba01442 Streptoco
23	364.8	5.2	22157	6	ABA01448	Aba01448 Streptoco
24	347.2	5.0	14142	6	ABA01435	Aba01435 Streptoco
25	346	4.9	12786	6	ABA01433	Aba01433 Streptoco
26	332	4.7	14460	6	ABA01440	Aba01440 Streptoco
27	320.2	4.6	1386	6	ABN70387	ABN70387 Streptoco
28	320.2	4.6	1386	6	ABN67429	ABN67429 Streptoco
29	314.8	4.5	15347	6	ABA01439	Aba01439 Streptoco
30	301.4	4.3	8056	8	ABZ10246	Abz10246 Haematopo
31	298.8	4.3	8056	8	ABZ10246	Abz10246 Haematopo
32	291.6	4.2	16110	6	ABA01437	Aba01437 Streptoco
33	283.8	4.1	1002	10	ADC50017	Adc50017 N-acetyl
34	255	3.6	447	6	ABN67430	ABN67430 Streptoco
35	230.2	3.3	12494	6	ABA01438	Aba01438 Streptoco
36	222.2	3.2	8056	8	ABZ10100	Abz10100 Haematopo
37	219.8	3.1	8056	8	ABZ10100	Abz10100 Haematopo
38	206.6	3.0	50000	6	ABL55643	AbL55643 AmEPV gen
39	198.4	2.8	15794	6	ABA01434	Aba01434 Streptoco
40	187.8	2.7	5979	4	AAS45313	Aas45313 Chemical
41	187.8	2.7	5979	6	ABK28152	ABK28152 DNA trans
42	185.8	2.7	50000	6	ABL55643	AbL55643 AmEPV gen
43	183.6	2.6	8170	6	ABK28258	ABK28258 DNA trans
44	179	2.6	11691	6	ABL34241	AbL34241 Human imm
45	176	2.5	372	8	ACA50223	ACA50223 Prokaryot

## ALIGNMENTS

RESULT 1  
AAZ60930  
ID AAZ60930 standard; DNA; 6992 BP.

AC AAZ60930;

DT 30-MAY-2000 (first entry)

DE Nucleotide sequence of a capsular gene cluster of S. suis serotype 1.

XX Capsular gene cluster; serotype 1; polysaccharide biosynthesis;  
XX capsular component; antigen; regulation; chain length determination;  
XX complement-mediated opsonophagocytosis; serotype-specific detection;  
XX antigen; vaccine; Streptococcal disease; CpsII; CpsI; CpsIG; CpsIH;  
XX CpsII; CpsIJ; CpsIK; glycosyltransferase; Cp polymerase; ss.

OS Streptococcus suis.

XX	Key	Location/Qualifiers
FT	FD	2..1367
FT	CDS	/*tag= a
FT		/note= "CpsIE; encodes AAY68970"
FT	CDS	1374..1823
FT		/*tag= b
FT		/note= "CpsIF; encodes AAY68971"
FT	CDS	1823..2317
FT		/*tag= c
FT		/note= "CpsIG; encodes AAY68972"
FT	CDS	3036..4202
FT		/*tag= d
FT		/note= "CpsIH; encodes AAY68973; the nucleotides encoding
FT		amino acid 213 are not given"
FT	CDS	4195..5163
FT		/*tag= e
FT		/note= "CpsII; encodes AAY68974"
FT	CDS	5172..6143
FT		/*tag= f
FT		/transl_except= (pos: 5634..5636, aa: Xaa)
FT		/transl_except= (pos: 5637..5638, aa: Xaa)
FT		/note= "CpsIJ; Xaa is an unspecified amino acid; encodes
FT		AAY68975; nucleotides 5328-5330 encode an amino acid that
FT		is not included in the protein sequence"
FT	CDS	6156..6990

```

FT      /*tag= g
FT      /note= "Cps1k; encodes AAY68976; no termination codon
FT      given"
XX      WO200005378-A2.
XX      03-FEB-2000.
XX      19-JUL-1999; 99WO-NL000460.
XX      22-JUL-1998; 98EP-00202465.
XX      22-JUL-1998; 98EP-00202467.
XX      (DIEN-) STICHTING DIENST LANDBOUWKUNDIG ONDERZOE.
XX      Smith HE;
XX      WPI; 2000-195104/17.
XX      P-PSDB; AAY68970, AAY68971, AAY68972, AAY68973, AAY68974, AAY68975,
XX      AAY68976.
XX      New nucleic acid containing the capsular gene cluster of Streptococcus
XX      suis, used for serotype-specific detection and to generate antigens or
XX      mutants for vaccination.
XX      Claim 5; Fig 4; 144pp; English.
XX      The present sequence represents the capsular gene cluster of
XX      Streptococcus suis serotype 1. The genes in this cluster are involved in
XX      polysaccharide biosynthesis of capsular components and antigens. The
XX      proteins have glycosyltransferase activities (CpsIE, CpsIG, CpsII, CpsIJ,
XX      CpsIK) and CP polymerase activities (CpsIH). The capsule confers
XX      bacterium resistance to complement-mediated opsonophagocytosis. The gene
XX      cluster is used as a source of probes and primers for serotype-specific
XX      detection of S. suis and is also useful for recombinant production of the
XX      proteins. The proteins are then useful for producing antigens that can be
XX      used in vaccines, for controlling or eradicating a Streptococcal disease,
XX      in humans or animals, e.g. against S. suis in pigs
XX      Sequence 6992 BP; 2513 A; 788 C; 1214 G; 2477 T; 0 U; 0 Other;
XX      Query Match 100.0%; Score 6992; DB 3; Length 6992;
XX      Best Local Similarity 100.0%; Pred. No. 0;
XX      Matches 6992; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 ATCCGCAAAAGCAATGGCATTATTTGATATGATAGCAGTTGCAATTTCTGCAATCTTAA 60
DB      1 ATCCGCAAAAGCAATGGCATTATTTGATATGATAGCAGTTGCAATTTCTGCAATCTTAA 60
QY      61 CAAGTCATATACCAAAATGCTGATTTAAATCGTTCTGGAAATTTTATCATATGATGTTTC 120
DB      61 CAAGTCATATACCAAAATGCTGATTTAAATCGTTCTGGAAATTTTATCATATGATGTTTC 120
QY      121 ATTTATTTTGGCAATTTTATATCTCGTATGCGAGTTGAATTTGAGTATAGAGGTAATCTGA 180
DB      121 ATTTATTTTGGCAATTTTATATCTCGTATGCGAGTTGAATTTGAGTATAGAGGTAATCTGA 180
QY      181 TAGAGTTTGGAAAAACATTTAACTATAGTATAATTTGCAATTTTCTTACGGCAGTAT 240
DB      181 TAGAGTTTGGAAAAACATTTAACTATAGTATAATTTGCAATTTTCTTACGGCAGTAT 240
QY      241 CATTTTGTGGAGAAATAATTCGCACTTTCAAGACGTTGCGCGTGTATTTTCAATTA 300
DB      241 CATTTTGTGGAGAAATAATTCGCACTTTCAAGACGTTGCGCGTGTATTTTCAATTA 300
QY      301 TAAACTTCGTTTGGTATACCTATTAACTGTAATTTATAGCAGTTTAAAGTAGCTTTC 360
DB      301 TAAACTTCGTTTGGTATACCTATTAACTGTAATTTATAGCAGTTTAAAGTAGCTTTC 360
QY      361 TATTTTTCGCAATCTATCAAAAAAGACGATTCTTAATTAACAACGGCTGAACGATGGGAAA 420
DB      361 TATTTTTCGCAATCTATCAAAAAAGACGATTCTTAATTAACAACGGCTGAACGATGGGAAA 420

```

```

QY      421 ATATGCAAGTTTATTTGATCACAATAAACAATTCAAAATAATCTTGTTCATTCGTAG 480
DB      421 ATATGCAAGTTTATTTGATCACAATAAACAATTCAAAATAATCTTGTTCATTCGTAG 480
QY      481 TTTTAGGTACAGAAATAGATAAAATTAATTTATCATTACCGCTCTATTTCTTGTGGAAG 540
DB      481 TTTTAGGTACAGAAATAGATAAAATTAATTTATCATTACCGCTCTATTTCTTGTGGAAG 540
QY      541 AAGCTATAGAGTTTCAACAAGGGAAGTGTGCAACACGTCCTTTATAAATCTACCAAGTG 600
DB      541 AAGCTATAGAGTTTCAACAAGGGAAGTGTGCAACACGTCCTTTATAAATCTACCAAGTG 600
QY      601 AGTTTGTAGACGTAAGCAATTCGTTTCAGATTTTGAATTTGAGTTTGTAGTATGATGAAGC 660
DB      601 AGTTTGTAGACGTAAGCAATTCGTTTCAGATTTTGAATTTGAGTTTGTAGTATGATGAAGC 660
QY      661 TTGATATTAATTCATTTCGGTTTCTGCTGCTTGAATAAACAATAAATTCCTAGTGTG 720
DB      661 TTGATATTAATTCATTTCGGTTTCTGCTGCTTGAATAAACAATAAATTCCTAGTGTG 720
QY      721 ACCATAGCATTTGAACTTTTCCACAAATTTTATAAGCCTAGTCATATCATGATGAAC 780
DB      721 ACCATAGCATTTGAACTTTTCCACAAATTTTATAAGCCTAGTCATATCATGATGAAC 780
QY      781 GACTTTTGGATATACCTCGGAGCGGTAGTCGGGTAAATTTTGTGGTATAGTTTCTATTT 840
DB      781 GACTTTTGGATATACCTCGGAGCGGTAGTCGGGTAAATTTTGTGGTATAGTTTCTATTT 840
QY      841 TGTAGTTTCAATTTATTCGTAGAGATGGTGACCGCTATTTTGTCTCAGAAAACGAGTTG 900
DB      841 TGTAGTTTCAATTTATTCGTAGAGATGGTGACCGCTATTTTGTCTCAGAAAACGAGTTG 900
QY      901 GACAGAATGGAAGCATATTTTACATTTCTCAAGTTTTCGATCGATGTATGTTGATGCTGAGG 960
DB      901 GACAGAATGGAAGCATATTTTACATTTCTCAAGTTTTCGATCGATGTATGTTGATGCTGAGG 960
QY      961 AGCGCAAAAAGACCTTGCTCAGCAAAAACAGATGCAAGGTGGTATGTTTTAAATGG 1020
DB      961 AGCGCAAAAAGACCTTGCTCAGCAAAAACAGATGCAAGGTGGTATGTTTTAAATGG 1020
QY      1021 GAAAAACGATCTAGAAATTAATTCCTCAATTTGACATTTTATACGCAAAAACAAAGTTTAGACG 1080
DB      1021 GAAAAACGATCTAGAAATTAATTCCTCAATTTGACATTTTATACGCAAAAACAAAGTTTAGACG 1080
QY      1081 AGTTACCACAGTTTATTAATTTTAAATTTGGCGATATGAGTCTAGTTGGTACAGCTCCAC 1140
DB      1081 AGTTACCACAGTTTATTAATTTTAAATTTGGCGATATGAGTCTAGTTGGTACAGCTCCAC 1140
QY      1141 CTACAGTTGATGAATTTGAAAAATATATCTCTCGTCAAAAAGACAGATTTGAGTTTAAAC 1200
DB      1141 CTACAGTTGATGAATTTGAAAAATATATCTCTCGTCAAAAAGACAGATTTGAGTTTAAAC 1200
QY      1201 CAGGGAATTAACAGTCTCTGCGAGTTAGTGTGCTAGTATATATACACAGCTTCGACGACG 1260
DB      1201 CAGGGAATTAACAGTCTCTGCGAGTTAGTGTGCTAGTATATATACACAGCTTCGACGACG 1260
QY      1261 TAGTTCGGTTGGACTTTAGCATACATTTGGAATTTGGACTATCTGTCAGATATTTAAATTT 1320
DB      1261 TAGTTCGGTTGGACTTTAGCATACATTTGGAATTTGGACTATCTGTCAGATATTTAAATTT 1320
QY      1321 TATTAAGACAGTGAAAGTTGTTTGTGAGAGAGGGAAGTAAAGTAAAGTATATGAAG 1380
DB      1321 TATTAAGACAGTGAAAGTTGTTTGTGAGAGAGGGAAGTAAAGTAAAGTATATGAAG 1380
QY      1381 TTTGTTTGGTTCGTTCTTACGGGGGACATTTGACTCATTGTTATTTGTTTAAACCGTTT 1440
DB      1381 TTTGTTTGGTTCGTTCTTACGGGGGACATTTGACTCATTGTTATTTGTTTAAACCGTTT 1440
QY      1441 GGAAGGAAGAAAGACGTTTTTTTGGGTAAACATTTGATAAAGAGGATGCAAGAGTCTTTTGA 1500
DB      1441 GGAAGGAAGAAAGACGTTTTTTTGGGTAAACATTTGATAAAGAGGATGCAAGAGTCTTTTGA 1500
QY      1501 AGAATGAAAAAATGTATCCTTACTTCCAAATAATCGCAATCTCATTAATTAAGTGA 1560

```



Qy	3721	TTGCTTGGATAAAGCTAATAGTAATATTTGTAATACTACTATTATTATTAAACTG	3780
Db	3721	TTGCTTGGATAAAGCTAATAGTAATATTTGTAATACTACTATTATTATTAAACTG	3780
Qy	3781	AAATTGCTTTACATGAATTTTGGCTGTTTATAATCTAGAGAAATCAAGTAACGAAGCTA	3840
Db	3781	AAATTGCTTTACATGAATTTTGGCTGTTTATAATCTAGAGAAATCAAGTAACGAAGCTA	3840
Qy	3841	GATTTATTTATTTACAGGAAGTATGATAAAGTATTAGAAAAACAATTTTATTTCGAT	3900
Db	3841	GATTTATTTATTTACAGGAAGTATGATAAAGTATTAGAAAAACAATTTTATTTCGAT	3900
Qy	3901	ATGGAATATCCGATATTCAGTTACGGGAATCTGGCTCGGAAGTCATTCAGGCTATATAT	3960
Db	3901	ATGGAATATCCGATATTCAGTTACGGGAATCTGGCTCGGAAGTCATTCAGGCTATATAT	3960
Qy	3961	CAATTTTTTATAAATCAGGAATAGTTGGTGTGATTTTACTGATGTTTTCTTTTTTATG	4020
Db	3961	CAATTTTTTATAAATCAGGAATAGTTGGTGTGATTTTACTGATGTTTTCTTTTTTATG	4020
Qy	4021	TTATAAAAAAGTTATGGAGTTAATGGGGAACAGCACTATTTTATTTACATCATTAG	4080
Db	4021	TTATAAAAAAGTTATGGAGTTAATGGGGAACAGCACTATTTTATTTACATCATTAG	4080
Qy	4081	CCATATTTTTCATATATGAACAATAGATCCGATATATATATATAGTACTATCTTTT	4140
Db	4081	CCATATTTTTCATATATGAACAATAGATCCGATATATATATATAGTACTATCTTTT	4140
Qy	4141	CTTCAATAGGTATTTGGAATAATAAATTTTAAAAAGGATATGAGACAAAAAATGAAT	4200
Db	4141	CTTCAATAGGTATTTGGAATAATAAATTTTAAAAAGGATATGAGACAAAAAATGAAT	4200
Qy	4201	GATTTAAATTTTCAGTTATTTGTAACCAATTTATAATGTCCTCAAGATTTATCTTGATAAATGATT	4260
Db	4201	GATTTAAATTTTCAGTTATTTGTAACCAATTTATAATGTCCTCAAGATTTATCTTGATAAATGATT	4260
Qy	4261	AACAGTATTTATACCAAAACATATACTAAATTTAGAGGTTATCTCGTAATATGATGGAAGT	4320
Db	4261	AACAGTATTTATACCAAAACATATACTAAATTTAGAGGTTATCTCGTAATATGATGGAAGT	4320
Qy	4321	ACTGATGATCTGAGAAAAATTTGCTTAACTATATGAAGAACGATGGAAGATTTAAATAT	4380
Db	4321	ACTGATGATCTGAGAAAAATTTGCTTAACTATATGAAGAACGATGGAAGATTTAAATAT	4380
Qy	4381	TACAAGAAAAATTAATGGCGGTCTAGCAGATGCTCGAAAATTTTCGGACTAGAACATGCAACA	4440
Db	4381	TACAAGAAAAATTAATGGCGGTCTAGCAGATGCTCGAAAATTTTCGGACTAGAACATGCAACA	4440
Qy	4441	GGTAAATATATTTGCTTTTGTGCGATTTCTGATGACTATATAGAAAGTTGCAATGTTTCGAGAGA	4500
Db	4441	GGTAAATATATTTGCTTTTGTGCGATTTCTGATGACTATATAGAAAGTTGCAATGTTTCGAGAGA	4500
Qy	4501	ATGCATGATATAATAAATCTGAGTATAATGCGCATATAGCAGAGATAGATTTTGTGTTAGTA	4560
Db	4501	ATGCATGATATAATAAATCTGAGTATAATGCGCATATAGCAGAGATAGATTTTGTGTTAGTA	4560
Qy	4561	GACGAAAAACGGGTATACAAAGAAAAAAGAAATAGTAAATTTTCATGTCCTTAACCGAGAGAA	4620
Db	4561	GACGAAAAACGGGTATACAAAGAAAAAAGAAATAGTAAATTTTCATGTCCTTAACCGAGAGAA	4620
Qy	4621	GAGACTGTAAAGAAATTTTGTCTAGGATCTAATATAGAAAAATATGTTGGTGGCAAGCTT	4680
Db	4621	GAGACTGTAAAGAAATTTTGTCTAGGATCTAATATAGAAAAATATGTTGGTGGCAAGCTT	4680
Qy	4681	TATTCACGAGATATTAATAAAGATATAAAATTCCAAATTTAATATAGAAATATGTTGGTGGAG	4740
Db	4681	TATTCACGAGATATTAATAAAGATATAAAATTCCAAATTTAATATAGAAATATGTTGGTGGAG	4740
Qy	4741	GATTTGCTTTTAAATTTGGAGGCTCTGAAACAATGTAAACACGCTGTAGTGTGACTAGTA	4800
Db	4741	GATTTGCTTTTAAATTTGGAGGCTCTTGAACAATGTAAACACGCTGTAGTGTGACTAGTA	4800

Qy	4801	GAATATTTATTAATTTATGTCATTCCTTAACAGATTCCCTTATTAATCAGAAATTTCTCTATA	4860
Db	4801	GAATATTTATTAATTTATGTCATTCCTTAACAGATTCCCTTATTAATCAGAAATTTCTCTATA	4860
Qy	4861	AATAATATTTAGTTAGTCAAGATTGGAGAAATTTAGCCCTTTTAAGTTAAAAAGAGAGTTT	4920
Db	4861	AATAATATTTAGTTAGTCAAGATTGGAGAAATTTAGCCCTTTTAAGTTAAAAAGAGAGTTT	4920
Qy	4921	AGTCATTTATTTGATCGAAAGATTATTAAGAGAGGTTTAAATGTTTAAACAAAAATGTAT	4980
Db	4921	AGTCATTTATTTGATCGAAAGATTATTAAGAGAGGTTTAAATGTTTAAACAAAAATGTAT	4980
Qy	4981	TCAACAGATTGTTTGGATATGATGTTCTTGCSCCAATATTAAGAGCTTATCCGAAAGAAATA	5040
Db	4981	TCAACAGATTGTTTGGATATGATGTTCTTGCSCCAATATTAAGAGCTTATCCGAAAGAAATA	5040
Qy	5041	CGTAGATATCCATTTTAAAGCGAAAGAGATATTTATCAAGAAAGCATTTTAGTTACGTTG	5100
Db	5041	CGTAGATATCCATTTTAAAGCGAAAGAGATATTTATCAAGAAAGCATTTTAGTTACGTTG	5100
Qy	5101	TATTCATGAAATTTTTCGCCCTAAACTATATGTAATGTTATATAGAAATTTCAAAAAGCAG	5160
Db	5101	TATTCATGAAATTTTTCGCCCTAAACTATATGTAATGTTATATAGAAATTTCAAAAAGCAG	5160
Qy	5161	TAGAGCTTAAAAATGATATAAAATTTAGTGTATTTCCAGTTTATTAATGTAAGATAATTT	5220
Db	5161	TAGAGCTTAAAAATGATATAAAATTTAGTGTATTTCCAGTTTATTAATGTAAGATAATTT	5220
Qy	5221	TAAAGTAGTTGTAAGAAAGCATTTTAAATCAAAATTTATAAAAATATAGAAATTTATTTGA	5280
Db	5221	TAAAGTAGTTGTAAGAAAGCATTTTAAATCAAAATTTATAAAAATATAGAAATTTATTTGA	5280
Qy	5281	TAGATGATGGCTCTGATAGATTTCTGCTAAATAATGCAAGAAATATGCAAGAAAGATA	5340
Db	5281	TAGATGATGGCTCTGATAGATTTCTGCTAAATAATGCAAGAAATATGCAAGAAAGATA	5340
Qy	5341	AAAGCTTAAAAATTTTTCACCTAAATCAGTAGCGAGTATCAAAATGCTAGAAATCATGGAA	5400
Db	5341	AAAGCTTAAAAATTTTTCACCTAAATCAGTAGCGAGTATCAAAATGCTAGAAATCATGGAA	5400
Qy	5401	TAAAGCGGAGTACAGCTGAATATTTATGTTGTTGCTGACTCTGATGATGTTTGTGATAGTA	5460
Db	5401	TAAAGCGGAGTACAGCTGAATATTTATGTTGTTGCTGACTCTGATGATGTTTGTGATAGTA	5460
Qy	5461	GATTAGTAGAAAAATTTATATTTTAAATATATAAAAAAGTAGAAGTATCTGGTTGTT	5520
Db	5461	GATTAGTAGAAAAATTTATATTTTAAATATATAAAAAAGTAGAAGTATCTGGTTGTT	5520
Qy	5521	TGTACGCTACTTTTTCAGAAAAATATAAATTTTGAAGTGAATTAATCCAAATATTCGATT	5580
Db	5521	TGTACGCTACTTTTTCAGAAAAATATAAATTTTGAAGTGAATTAATCCAAATATTCGATT	5580
Qy	5581	TTCAAGCAATTAATACCGTGCAGGACATGGGAGAAAAAATTTTATGAAATTTGTATATAA	5640
Db	5581	TTCAAGCAATTAATACCGTGCAGGACATGGGAGAAAAAATTTTATGAAATTTGTATATAA	5640
Qy	5641	ATAATATTTTCTACTCTCTGTTTAACTATAATAAGAAAGATACATAACAGATCTTT	5700
Db	5641	ATAATATTTTCTACTCTCTGTTTAACTATAATAAGAAAGATACATAACAGATCTTT	5700
Qy	5701	TTCAAGAGATCAATGTTAGGAGAGATTTTATCTTTTAACTCTGCAATTTTAAAGAAATA	5760
Db	5701	TTCAAGAGATCAATGTTAGGAGAGATTTTATCTTTTAACTCTGCAATTTTAAAGAAATA	5760
Qy	5761	TAGATAGAGTTAGTTATTTGACTGAACATCTTTTATTTTATAGGAGAGGTATATAAGTA	5820
Db	5761	TAGATAGAGTTAGTTATTTGACTGAACATCTTTTATTTTATAGGAGAGGTATATAAGTA	5820
Qy	5821	CAGTAAATCTTTTAAAGAGGTTGTTTGTCAATTTGGAAAAATTTTGCACAAAAACAAGTGA	5880
Db	5821	CAGTAAATCTTTTAAAGAGGTTGTTTGTCAATTTGGAAAAATTTTGCACAAAAACAAGTGA	5880
Qy	5881	TAGTATTTGTTTAAAGCAAAATATATGTTGAGGATTTTTCAGCTGATCAATTTGTTTAAAGATACTA	5940

Db 5881 TAGTATTGTTTAAAGCAATATATGTTGAGGATTTTGGCGTATCAATTTGTTAAAGATACTA 5940  
Qy 5941 TACGTTGGCAAGTATTTTATTATAGCTTACTAAATGTTTAAATACGGAACACAGTCTATTT 6000  
Db 5941 TACGTTGGCAAGTATTTTATTATAGCTTACTAAATGTTTAAATACGGAACACAGTCTATTT 6000  
Qy 6001 TTGACAAATTTTAAATTTTAAAGAAATCTTTTATAAAATATATTTTAACTTTGTTAAAG 6060  
Db 6001 TTGACAAATTTTAAATTTTAAAGAAATCTTTTATAAAATATATTTTAACTTTGTTAAAG 6060  
Qy 6061 TATCTACAAATTTCTTTGCTAAATTTTCTTAAAGAAATTTGTTTAAAGAAATTTTCTGACAAAGTTT 6120  
Db 6061 TATCTACAAATTTCTTTGCTAAATTTTCTTAAAGAAATTTGTTTAAAGAAATTTTCTGACAAAGTTT 6120  
Qy 6121 TTTAAAAAATATTATGTTTATAATAGGAAGATATCATGATCACTATTAGTAAATTTCTA 6180  
Db 6121 TTTAAAAAATATTATGTTTATAATAGGAAGATATCATGATCACTATTAGTAAATTTCTA 6180  
Qy 6181 TAAATGTACCTATATATTAATCTAGAAAATATTTATCTAAATGTATAGATGACATTTGTA 6240  
Db 6181 TAAATGTACCTATATATTAATCTAGAAAATATTTATCTAAATGTATAGATGACATTTGTA 6240  
Qy 6241 ATCAGACCTACAAACATATAGAGATTTCTTCTGGTGAATGACGGTAGTACGATTAATTCGG 6300  
Db 6241 ATCAGACCTACAAACATATAGAGATTTCTTCTGGTGAATGACGGTAGTACGATTAATTCGG 6300  
Qy 6301 AAGAAATTTGTTTAGCATATCGAAGAAAGATAGTCGCATTCGTTATTTTAAAAAAGAGA 6360  
Db 6301 AAGAAATTTGTTTAGCATATCGAAGAAAGATAGTCGCATTCGTTATTTTAAAAAAGAGA 6360  
Qy 6361 ACGGCGGCTATCAGATGCCGCTAATTAATGATGATAGTAAAGTCGCGCAAGGGTACACTTAG 6420  
Db 6361 ACGGCGGCTATCAGATGCCGCTAATTAATGATGATAGTAAAGTCGCGCAAGGGTACACTTAG 6420  
Qy 6421 CTTTTATAGACTCAGATGATTTTATTCATTCGGAGTTTCATCCAGGTTTACACGAACAA 6480  
Db 6421 CTTTTATAGACTCAGATGATTTTATTCATTCGGAGTTTCATCCAGGTTTACACGAACAA 6480  
Qy 6481 TTGAGAGAGAAATGCCCTGTCGCGAGTTGCTGTTATGATAGGTTAGATGCTTCGGGCG 6540  
Db 6481 TTGAGAGAGAAATGCCCTGTCGCGAGTTGCTGTTATGATAGGTTAGATGCTTCGGGCG 6540  
Qy 6541 ATTTCTTAAACAGCAGACGCGCTTCTCAAAATCAGGCTGTTCTGAGCGGCAAGGAATGTTT 6600  
Db 6541 ATTTCTTAAACAGCAGACGCGCTTCTCAAAATCAGGCTGTTCTGAGCGGCAAGGAATGTTT 6600  
Qy 6601 GTAAAAAGCTCTAGAGCGGATGTCATCGCTTTGTTGTCGCTGTATTAATTAACCTCTATA 6660  
Db 6601 GTAAAAAGCTCTAGAGCGGATGTCATCGCTTTGTTGTCGCTGTATTAATTAACCTCTATA 6660  
Qy 6661 AAAAAGAACTATTTGAAGATTTTCGATTTGAAAAGGGTAAAGTTCATGAAGATGAATACT 6720  
Db 6661 AAAAAGAACTATTTGAAGATTTTCGATTTGAAAAGGGTAAAGTTCATGAAGATGAATACT 6720  
Qy 6721 TCACCTATCGCTGCTCTATCAGTTAGAAAAGTTGCAATAGTTAAGGAGTGTGTTGACT 6780  
Db 6721 TCACCTATCGCTGCTCTATCAGTTAGAAAAGTTGCAATAGTTAAGGAGTGTGTTGACT 6780  
Qy 6781 ATTATGTTGACCGAATAATAGTATCAAACTTCTAGCATCACTGACCATCGCTTCCATT 6840  
Db 6781 ATTATGTTGACCGAATAATAGTATCAAACTTCTAGCATCACTGACCATCGCTTCCATT 6840  
Qy 6841 GCCTACTGGATTTCAAAATGACAAATGGAATCTTATGAAGTAGAGGATTAAGAGC 6900  
Db 6841 GCCTACTGGATTTCAAAATGACAAATGGAATCTTATGAAGTAGAGGATTAAGAGC 6900  
Qy 6901 TCTTACTAGAGTGTATCGTTTCATTTTATGCTTTGCTTTGTTTTAGGCAAAATATA 6960  
Db 6901 TCTTACTAGAGTGTATCGTTTCATTTTATGCTTTGCTTTGTTTTAGGCAAAATATA 6960  
Qy 6961 ATCATTTGTTGAGCAACACGAAAGAGCTT 6992

Db 6961 ATCATTTGTTGAGCAACACGAAAGAGCTT 6992  
RESULT 2  
ID AA260929 standard; DNA; 26281 BP.  
XX AA260929;  
AC AA260929;  
XX 30-MAY-2000 (first entry)  
DT  
XX Nucleotide sequence of a capsular gene cluster of *S. suis* serotype 2.  
DE  
XX Capsular gene cluster; serotype 2; polysaccharide biosynthesis;  
KW capsular component; antigen; regulation; chain length determination;  
KW complement-mediated opsonophagocytosis; serotype-specific detection;  
KW antigen; vaccine; Streptococcal disease; ORF 22; ORF 2Y; ORF 2Z; Cps2A;  
KW Cps2B; Cps2C; Cps2D; Cps2E; Cps2F; Cps2G; Cps2H; Cps2I; Cps2J; Cps2K;  
KW Cps2O; Cps2P; Cps2Q; Cps2R; Cps2S; Cps2T; ss.  
XX Streptococcus suis.  
OS  
XX  
FH Key Location/Qualifiers  
CDS 2..721  
FT /\*tag= a  
FT /note= "ORF 22; encodes AAY68950"  
FT complement (822..2079)  
FT /\*tag= b  
FT /note= "ORF 2Y; encodes AAY68951"  
FT 2202..2936  
FT /\*tag= c  
FT /transl\_except= (pos: 2916..2918, aa: Xaa)  
FT /note= "ORF 2X; Xaa is an unspecified amino acid; encodes  
FT AAY68952"  
FT 3041..4486  
FT /\*tag= d  
FT /note= "Cps2A; encodes AAY68953"  
FT 4504..5193  
FT /\*tag= e  
FT /note= "Cps2B; encodes AAY68954"  
FT 5203..5880  
FT /\*tag= f  
FT /note= "Cps2C; encodes AAY68955"  
FT 5919..6650  
FT /\*tag= g  
FT /note= "Cps2D; encodes AAY68956"  
FT 6675..8054  
FT /\*tag= h  
FT /note= "Cps2E; encodes AAY68957"  
FT 8089..9258  
FT /\*tag= i  
FT /note= "Cps2F; encodes AAY68958"  
FT 9262..10419  
FT /\*tag= j  
FT /transl\_except= (pos: 10057..10059, aa: Xaa)  
FT /note= "Cps2G; Xaa is an unspecified amino acid; encodes  
FT AAY68959"  
FT 10808..12176  
FT /\*tag= k  
FT /transl\_except= (pos: 11963..11965, aa: Xaa)  
FT /note= "Cps2H; Xaa is an unspecified amino acid; encodes  
FT AAY68960"  
FT 12213..13445  
FT /\*tag= l  
FT /note= "Cps2I; encodes AAY68961"  
FT 13583..14581  
FT /\*tag= m  
FT /note= "Cps2J; encodes AAY68962"  
FT 14574..15578  
FT /\*tag= n  
FT /note= "Cps2K; encodes AAY68963"  
FT 18401..18904  
FT /\*tag= o  
FT

```
FT /transl_except= (pos: 18755..18957, aa: Ile)
FT /transl_except= (pos: 18770..18772, aa: Leu)
FT /transl_except= (pos: 18777..18778, aa: Ile)
FT /transl_except= (pos: 18836..18838, aa: Pro)
FT /transl_except= (pos: 18890..18892, aa: Thr)
FT /transl_except= (pos: 18896..18898, aa: Pro)
FT /transl_except= (pos: 18935..18937, aa: Pro)
FT /transl_except= (pos: 18953..18955, aa: Leu)
FT /transl_except= (pos: 18968..18970, aa: Xaa)
FT /transl_except= (pos: 18971..18973, aa: Trp)
FT /transl_except= (pos: 18977..18979, aa: Ser)
FT /transl_except= (pos: 18980..18982, aa: Gln)
FT /transl_except= (pos: 18989..18991, aa: Asn)
FT /note= "Cpe20; Xaa is an unspecified amino acid; encodes
FT AAY68964"
FT CDS 20327..21343
FT /*tag= p
FT /note= "Cpe2P; encodes AAY68965"
FT CDS 21355..21867
FT /*tag= q
FT /note= "Cpe2Q; encodes AAY68966"
FT CDS 21933..22487
FT /*tag= x
FT /note= "Cpe2R; encodes AAY68967"
FT CDS 22501..23127
FT /*tag= s
FT /note= "Cpe2S; encodes AAY68968"
FT CDS 23136..24368
FT /*tag= t
FT /note= "Cpe2T; encodes AAY68969"
FT XX
FT WO200005378-A2.
FT XX
FT PD 03-FEB-2000.
FT XX
FT PD 19-JUL-1999; 99WO-NL000460.
FT PF
FT PR 22-JUL-1998; 98EP-00202465.
FT PR 22-JUL-1998; 98EP-00202467.
FT XX
FT PA (DIEN-) STICHTING DIENST LANDBOUWKUNDIG ONDERZOE.
FT XX
FT DI Smith HE;
FT XX
FT DR WPI; 2000-195104/17.
FT DR P-PSDB; AAY68950, AAY68951, AAY68952, AAY68954, AAY68955,
FT DR AAY68956, AAY68957, AAY68958, AAY68959, AAY68960, AAY68961, AAY68962,
FT DR AAY68963, AAY68964, AAY68965, AAY68966, AAY68967, AAY68968, AAY68969.
FT XX
FT PT New nucleic acid containing the capsular gene cluster of Streptococcus
FT PT suis, used for serotype-specific detection and to generate antigens or
FT PT mutants for vaccination.
FT XX
FT PS Claim 4; Fig 3; 144pp; English.
FT XX
FT CC The present sequence represents the capsular gene cluster of
FT CC Streptococcus suis serotype 2. The genes in this cluster are involved in
FT CC polysaccharide biosynthesis of capsular components and antigens. The
FT CC proteins are involved in regulation (CpsA), chain length determination
FT CC (CpsB, CpsC), export (CpsC), and biosynthesis (CpsE, CpsF, CpsG, CpsH,
FT CC CpsJ, CpsK). The capsule confers bacterium resistance to complement-
FT CC mediated opsonophagocytosis. The gene cluster is used as a source of
FT CC probes and primers for serotype-specific detection of S. suis and is also
FT CC useful for recombinant production of the proteins. The proteins are then
FT CC useful for producing antigens that can be used in vaccines, for
FT CC controlling or eradicating a Streptococcal disease, in humans or animals,
FT CC e.g. against S. suis in pigs
FT XX
FT SQ Sequence 26281 BP; 8252 A; 4158 C; 5488 G; 8383 T; 0 U; 0 Other;
FT Query Match 17.1%; Score 1195.8; DB 3; Length 26281;
FT Best Local Similarity 78.9%; Pred. No. 1.1e-151;
FT Matches 1437; Conservative 0; Mismatches 382; Indels 2; Gaps 1;
```

```
Qy 5172 ATCGATAAAATTAGTGTATTGTTCCAGTTTATAAATGTAGATAAAATATTTAAGTAGTTGT 5231
Db 13583 ATGGAATAATTCAGCATTTATTGTACCTATTTTTTAATACGGAAGTACTTTAAGAGAGTGT 13642
Qy 5232 ATAGAAGCATTTAATCAAAATTTAAAAATATAGAAATATATTATTGATAGATGATGGC 5291
Db 13643 TTAGATAGCATTTATTTCCCAATCGTATACTAATCTAGAGATTCTTTTGATAGATGACGGT 13702
Qy 5292 TCTGTAGATGATTCTGCTAAAAATATCAAGGAATATATCGAAAAAAGATAAAGAGTAAAA 5351
Db 13703 TCTTCAGATTCAACAACGGATATATGTTGGATACGACAGCAGATGCTAGATAAAA 13762
Qy 5352 ATTTTTTTCATTAATCATAGTGGAGTATCAAAATGCTAGAAATCATGGAATAAAGCGGAGT 5411
Db 13763 CTTTTCGGTTTACCAATGGTGGTGTTCAAAACGCAAGGAATTTACGGTATCAAAAATAGC 13822
Qy 5412 ACAGCTGAATATATTATGTTGTTGACTCTGATGATGTTGTTGATAGTAGTAGTAGAA 5471
Db 13823 ACAGCAAAATTAATTAATGTTGTAGATTCTGATGATATTGTTGACGGCAACATTTGTTGAG 13882
Qy 5472 AAATTATATTTTAAATATATAAAGTGAAGTGATTATCTGCTGTTGTTGTTGATCGTACT 5531
Db 13883 TCCTTATACACCTGTTTAAAGAGATGATAGTGAATTTGTCGGAGGGTTACTTGTCTACT 13942
Qy 5532 TTTTCAGAAAATATAAATAATTTTGAAGTGAATAATCCAAATATTTGATTTTGAAGCAATT 5591
Db 13943 TTTGATGGAAATTTATCAAGAAATCTGAGCTGCAAAAGTGTCAAAATTTGATTTGGAAGAGATA 14002
Qy 5592 AATACCGTCAGGACATGGGAGAAAATAATTTTATGAAATTTGATATATAAATAATATTTT 5651
Db 14003 AAAGAGTGGAGACTTAGGAAATGAAAATTTTCCCAATCATTTATATGACGGGTATCTTT 14062
Qy 5652 TCTACTCCTGTTGTTAAACTATATAAGAAAAGATACATACAGATCTTTTTCGAAGAAAT 5711
Db 14063 AATAGCCCTGTTGCAAACTTTATAGATATATATATATAACCAAGTTTTCACATGAA 14122
Qy 5712 CAATGTTAGGAGAAGATTTCATTTTAACTGCAATTTTAAAGAAATATAGATAGAGTT 5771
Db 14123 CAGTGGTTAGGAGAGGACTTATTATTATTAATCTAAATTTTAAAGAAATATAAAAAAGTC 14182
Qy 5772 AGTTATTTGACTGAACATCTTTATTTTATAGAGAGGATATACATAGTACAGTAAATCT 5831
Db 14183 CGCTATGTTAAACAGAAATCTTTATTTTGCAGAAAGAGTTTACAAAGTACTACAAATACG 14242
Qy 5832 TTTAAAGAAAGGTGTTGTTTTCCAATTTGCAAAATTTGCAAAAAAAGTGAATGATTTCTTT 5891
Db 14243 TTTAATAATGATGTTTATTCAATTTAGAAAATTTAGAGAAAACCTTTTGTATTTCTTT 14302
Qy 5892 AAGCAAAATATCGTGAGGATTTTGACGTATCAATTTGTTTAAAGATCTATACGTTGGCAA 5951
Db 14303 GTTAAATAATTTGGTGGCAATATGAATTTTCTGTTTAAAGAGACGCTACAGTGGCAT 14362
Qy 5952 GTATTTTATATAGCTTACTTAATGTTTAAATACGAAAAACAGTCTATTTTTCAGAAATTT 6011
Db 14363 ATTATTTATATAGCTTATTAAATGTTTCAAAAATGGAGATGAATCGCTTCCAAAGAAATG 14422
Qy 6012 TTAATTTTGAAGAAATCTTTTATAAAAAATATTTATTTTAACTTTGTTTAAAGTATCTAAACAA 6071
Db 14423 CATATATTT--AGTATTTATACAATAGGCAATCTTTAGATACCTCAAGTATTTAAACGAA 14480
Qy 6072 AATCTTTGTTAAAAATTTTGTATAAGAAATTTGTTTCGAACAAAGTTTAAAAAATA 6131
Db 14481 CGTCCCTCTGTTTAAAAAGAAATATGAAATTTAAATTTGTTGCTAATAATTTGTTAAAAATTT 14540
Qy 6132 TTATGGTTAATAAGGAGATATCATGATCTATTAGTAAATTTCTATAATTTGTACCT 6191
Db 14541 TTTTAAATCTTTAAATTTAGGGAAGAAAAAATTAATGATTAACATTTCTATCATCGTCCCA 14600
Qy 6192 ATATATAATGTAGAAAAATATTTTATCTAAATGTATAGATAGCATTTGTAATTCAGACCTAC 6251
Db 14601 ATTTACAATGTTGAACAATATCTATCCAAGTGTATAAATAGCATTTGTAATTCAGACCTAC 14660
```



```
QY 6252 AAACATATAGAGATCTTCTCGTGAATGACGGTAGTACGGATAAATTCGGAAGAAATTTGT 6311
Db 14661 AAACATATAGAGATCTTCTCGTGAATGACGGTAGTACGGATAAATTCGGAAGAAATTTGT 14720
QY 6312 TTAGCATATGCGAAGAAAGATAGTCGCAATTCGTTATTTTAAAGAAAGAAACGGCGGGCTA 6371
Db 14721 TTAGCATATGCGAAGAAAGATAGTCGCAATTCGTTATTTTAAAGAAAGAAACGGCGGGCTA 14780
QY 6372 TCAGATGCCGTAATTTATGGCATAAGTCGCGCCAGGGTGACTACTTAGCTTTTATATAGAC 6431
Db 14781 TCAGATGCCGTAATTTATGGCATAAGTCGCGCCAGGGTGACTACTTAGCTTTTATATAGAC 14840
QY 6432 TCAGATGATTTATTCATTCGAGTTTCATCAAGTTTACAGAGCAATTCGAGAGAG 6491
Db 14841 TCAGATGATTTATTCATTCGAGTTTCATCAAGTTTACAGAGCAATTCGAGAGAG 14900
QY 6492 AATGCCCTTGTGGCAGTTGCTGTTATCATAGGTAGATGCTTCGGGCAATTTCTTAACA 6551
Db 14901 AATGCCCTTGTGGCAGTTGCTGTTATCATAGGTAGATGCTTCGGGCAATTTCTTAACA 14960
QY 6552 GCAGAGCGCTTCTTACAAATCAGGCTGTTCTGAGCGCAGGAATGTTTGTAAAGAGCTG 6611
Db 14961 GCAGAGCGCTTCTTACAAATCAGGCTGTTCTGAGCGCAGGAATGTTTGTAAAGAGCTG 15020
QY 6612 CTAGAGCGGATGTCATCGCTTTGTGTCGCTGTATAAATCTATAAAGAACTA 6671
Db 15021 CTAGAGCGGATGTCATCGCTTTGTGTCGCTGTATAAATCTATAAAGAACTA 15080
QY 6672 TTTGAGATTTTCGATTTTGAAGGTTAAGATTCATGAGATGATGATCTTCACTTATCGC 6731
Db 15081 TTTGAGATTTTCGATTTTGAAGGTTAAGATTCATGAGATGATGATCTTCACTTATCGC 15140
QY 6732 TTGCTCTATGATGTAGAAAAGTTGCAATAGTAAAGAGTGCTGTACTATTATGTTGAC 6791
Db 15141 TTGCTCTATGATGTAGAAAAGTTGCAATAGTAAAGAGTGCTGTACTATTATGTTGAC 15200
QY 6792 CGAGAAATAGTATCACAACTTCTAGCATGACTGACCATCGCTTCCATTCGCTACTGAA 6851
Db 15201 CGAGAAATAGTATCACAACTTCTAGCATGACTGACCATCGCTTCCATTCGCTACTGAA 15260
QY 6852 TTTCAAATGACGATGACTTCTATGAAAGTAGAGAGATAAGAGCTCTTACTAGAG 6911
Db 15261 TTTCAAATGACGATGACTTCTATGAAAGTAGAGAGATAAGAGCTCTTACTAGAG 15320
QY 6912 TGTATCGTTCATTTTATGCTGCTGCTTTGTTGTTTATGAGCAATATAATCAATGTTGTTG 6971
Db 15321 TGTATCGTTCATTTTATGCTGCTGCTTTGTTGTTTATGAGCAATATAATCAATGTTGTTG 15380
QY 6972 AGCAAAACAGCAAAAGAGCTT 6992
Db 15381 AGCAAAACAGCAAAAGAGCTT 15401
```

## RESULT 3

```
ABA01441
ID ABA01441 standard; DNA; 17468 BP.
XX AC ABA01441;
XX AC ABA01441;
XX DT 21-FEB-2002 (first entry)
XX DE Streptococcus thermophilus eps3 operon #2.
XX KW Exopolysaccharide; lactic acid bacterium; eps; fermented food product;
XX KW ds.
XX OS Streptococcus thermophilus.
XX PN W0200179500-A2.
XX PD 25-OCT-2001.
XX PF 18-APR-2001; 2001WO-FR001199.
```

```
XX 18-APR-2000; 2000FR-00004972.
XX PR (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX PA (DANO-) CIE DANONE SA GERVAIS.
XX PA (RHOD ) RHODIA CHIM.
XX FI Rallu F, Besancon-Yoshpe I, Fremaux C, Mengaud J, Renault P;
XX WPI; 2002-017616/02.
XX DR
XX PT New nucleic acid fragments containing exopolysaccharide operon, useful
XX e.g. for increasing exopolysaccharide synthesis in lactic acid bacteria.
XX PS Claim 9; Page 75-80; 144pp; French.
XX CC The present sequence is an eps operon from Streptococcus thermophilus.
XX CC Proteins encoded by the eps operon function in exopolysaccharide (EPS)
XX CC synthesis. The operon is useful for producing chimeric eps operons, for
XX CC optimising production of EPS in lactic acid bacteria. EPS impart texture,
XX CC mouth feel and rheological properties to fermented food products (e.g.
XX CC yoghurt). They function as thickeners, to provide free-flowing and creamy
XX CC texture, and may also have biological activities beneficial to health
XX SQ Sequence 17468 BP; 5871 A; 2455 C; 3307 G; 5835 T; 0 U; 0 Other;
Query Match 13.7%; Score 955.6; DB 6; Length 17468;
Best Local Similarity 65.8%; Pred. No. 1.9e-119;
Matches 1484; Conservative 0; Mismatches 729; Indels 41; Gaps 5;
QY 1 ATCCCAAAAGCAATTTGCAATTTATGATAGTAGCAATTTCTGCAATTTCTTAA 60
Db 6294 ACCGCAATTTAGACTGGCGCTTTTAGAATTTGATTTGGGTAGTAGTGTCTCTCAATTA 6353
QY 61 CAAGTCATATACCAAAATGCTCAATTTAAATCGTTCTGGATTTTATCATATATGATGTTTC 120
Db 6354 CCACCCACTTGGCCACCCTCAAGATGAATAGTAGGGACGCTATTGATTTATTTGCCCTGC 6413
QY 121 ATATTTTGTCAATTTTATATCTCGTAGTCAGTTGAATTTTGTAGTATAGAGTAACTCTGA 180
Db 6414 ATTTTGTGCTTTTATTTTCTCAGTATGTCATCAGAAATTTTGTAGTAGAGCTACTTAA 6473
QY 181 TAGAGTTTGAATAAATTTTAACTATATAGTATAATTTTGAATTTTCTTACGGCAGTAT 240
Db 6474 TAGAGTTTGAAGCGTCTGCTAGGTATGTTTGAATTTTCTTCAATTTCTCTCATTTT 6533
QY 241 CATTTTGTGGAGAAATTTTCGACTTTCAAGACGTTGGCGGTGATTTTTCATTTAA 300
Db 6534 CTTTTATTTGGGAGATAGTTTATCATTTACGCGACGTTGGCATTTATTTACCATTC 6593
QY 301 TAAACTTTCGTTTTCGTATACCTATTATTAACGTAATTTATTAAGCAGTTTAAAGATAGCTTTC 360
Db 6594 TTACCTTTATCATCGTTTATATTAATCACTGCATGTTCTGTAGATACAAATACCTTCTTCT 6653
QY 361 TATTTTCGCAATCTATCAAAAAAGAGCAATTTCTAATTTACAAACGCTGAAACGATGGAAA 420
Db 6654 TTATGACTGCTGAACAACAGAGAAATCTTTTAGTCATTTACCACACTACTGAGAGATTCGAT 6713
QY 421 ATATGCAAGTTTATTTGAATCATAAAACAAATTTCAAAAAAATCTTGTTCGATTTGGTAG 480
Db 6714 TAATGGAGGGGCTCTTTGAATCGGATCAACTTCTCCATAATATTGCGGGTATTGTAG 6773
QY 481 TTTTAGGTACAGAAATAGATAAATTAATTTATCATTTACCGCTCTATTATTTCTGTGGAAG 540
Db 6774 TCATTTGG---AGACGGGTAGTGTGCGCTTTCAGAGGGAGTTCCAGTAGTGTCTTTTGATT 6830
QY 541 AAGCTATAGAGTTTTCACAAAGGAGTGGTTCGACACAGCTCTTTTATATAATCTTACCAAGTG 600
Db 6831 ATGCAATAGAGTTTTCGACCCCATGAGTTTTCGACCATGTTTATCACTTACCTACCGAGTG 6890
QY 601 AGTTTTTTAGAGCTAAAGCAATTCGTTTTCAGATTTTGTAGTTTGTAGTATTCATGTAACG 660
Db 6891 AACATTACGATCTCAAAACATCTTGTTCGATTTTGAAGTCAATGGGTATTGATGTGAGTG 6950
```

```
Qy 661 TTGATATTAACTTCATTCGGTTTACTCGTGTGAACAAACAAACAAATCCAACTGCTAGGTG 720
Db 6951 TAGATATTAAACACATCGATTCGTTTAAAGGCTTTAAACAAATAAAAAATCAACAGGTGGAG 7010
Qy 721 ACCATAGCATTTGTAACCTTTTCCACAAATTTTATAAGCCCTAGTCATATCATGATGAAC 780
Db 7011 ACCATAGCATCGTGACCTTTAACTCCAAATTTACTACAAACATAGCCATATCTTCTAAGC 7070
Qy 781 GACTTTTGGATATACCTCGAGCGGTAGTCGGGTTAAATTTATTTGGGTATAGTTCTATTT 840
Db 7071 GCATGTTGGATATCTTTGGGCTTTGATTTGCTCTTAAATTTGGGTCTGCTGGGATTTG 7130
Qy 841 TGTTAGTTCCAAATTTATTCGTAGATCGGTGACCGCTATTTTTCCTCAGAAACGAGTTG 900
Db 7131 TCTTAGCCCCCATCAITTCGGAAGATGGAGACCGGCCATTTTGTTCAGAAACGAGTAG 7190
Qy 901 GACAGAAATGACGCATATTTACATTTCTACAAGTTTCGATCGATGATGTTGATGCTCAGG 960
Db 7191 GGAATAACGGACGTATCTTTAAATTTTAAATTCGGTTCTATGTACACTGATGCGGAGG 7250
Qy 961 AGCCCAAAAAAGACTTGTCTCAGCCAAAACCAAGATGCAAGGGTGGGTATGTTTAAAAATGG 1020
Db 7251 A---AAGAAAGAAATGAAAGCACAATAATCAGATGCA-----CGGTGGAATGTTCAAGAT 7301
Qy 1021 GAAAAACGATCCTAGAAATTTACTCAATTTGACATTTTCATAGCCAAAAACAAAGTTAGACG 1080
Db 7302 GAATAACGATCCAGTATCACTCTATCGGTCGTTTATTCG-AAAAACGAGTCTGGACG 7360
Qy 1081 AGTTACACACAGTTTATAATGTTTTTAATTTGGCGATATGAGTCTAGTTGGTACACGTCAC 1140
Db 7361 AACTGCTCAGTTTACATATGTTGGTTCGGCATATGAGTCTTTGGGACTCGCCAC 7420
Qy 1141 CTACAGTTGATGAATTTGAAAAATATATCTCTGCTCAAAAGACAGATGAGTTTAAAC 1200
Db 7421 CAACTGTCGATGAATTTGAAAAATATACACCAAGTCAAAACGTCGCTCAGCTTCAAGC 7480
Qy 1201 CAGGGATTACAGGCTCTGGCAGGTTAGTGTCTAGTATATATACAGACTTCGACGACG 1260
Db 7481 CAGGAATACAGGCTCTATGGCAGCAAGTGTGCTGATTAATATACCAACTTTGATGAGG 7540
Qy 1261 TAGTTCGGTTGGACTTAGCATATGATTAATTCGACTATCTGCTCAGATATATAAAATTT 1320
Db 7541 TGGTAAGACTCGATGTTGAATATATCGATATCTGCTGATTTGGTTCGATATCAAGATAT 7600
Qy 1321 TATTAAGACAGTGAAGTTGTATTTGTTGAGAGGGAAGTAAGT----- 1365
Db 7601 TGTTGAAGACTATTGTCGTGTTTAAAAAAGAGGGAAGTAGGTAGAGCCCCCTCTTCA 7660
Qy 1366 -----AAAAATATGAAAGTTGTATTTGTTGAGAGGGAAGTAAGT----- 1365
Db 7661 TTCTTGTATTGGGAGAAAAAATTTTGTATTAGTGGTCACTCGAGGGGCAATTTA 7720
Qy 1413 ACTCACTGTATTGTTTAAACCGTTTGGAGGAAGAACGCTTTTGGGTAAACATTT 1472
Db 7721 GCACATTTAAATATGTTTAAACCAATCTGGAGAGATGAGAACGTTTGGGTAAACCTTT 7780
Qy 1473 GATAAAGAGGATGCAAGAAGCTTTTGAAGAAATGAAAAATGTATCCATGTACTTTCCA 1532
Db 7781 GATAAAGAGGATGCAAGAAGCATCTTAAAGATGAAAAATGTATTTCTGTCACTTTCT 7840
Qy 1533 ACAATCGCATCTCAATTAATTTAGTGAATAACTTTCTAGCTTTCAAAATTTTACGT 1592
Db 7841 ACAATCGAACTTTAAAAAATTTAATTAATAATACTTTTCTAGCTATTAAGTATTAATA 7900
Qy 1593 GATGAGAAACAGATGTTATTTATTTCTGTCGCGCGGTGTGTCCTCTTTTAC 1652
Db 7901 AAAGAAAAACCTGATGTCATTTATTTCTGAGACTGGGTAGCCGTTCCATCTTTTAC 7960
Qy 1653 ATCGGAAACATTTTGGAGCAAGACGATTTATATTGAAGTATTTGATCGAGTTTAAATA 1712
Db 7961 ATTCGTAACATTTTGGAGCTTAAGACGGTTTATATTGAAGTCTTTGTATAGAAATGATAAA 8020
```

```
Qy 1713 TCTACATTAACTGAAAACTAGTTTATCCGTAACAGATATTTTATTGTTCTAGTGGAA 1772
Db 8021 CCAACAGTTTACAGAAAAATTAGTGTACCTGTGACGGATAGTTCATTTGTTCAATGGAA 8080
Qy 1773 GAAATGAAGAAGGTATATCTTAATCTATTAACCTTGGGGAGTATTTTTTAATGATTTTTG 1832
Db 8081 GAGATGAAAAAGCTTTATCCAAAAGCTATTAATTTAGGGAGTATCTTTTAATGATTTTTG 8140
Qy 1833 TAAACAGTAGGAATCATGAACAAACAGTTTAATCGATGATGAAGAGATGATTTATTGA 1892
Db 8141 TAAACAGTAGGAACACACGAAACAACTTTTAAATCGCTTAATTCAGGAAGTGGATCATTTAG 8200
Qy 1893 AAAAAATGGAAGTATAAACCGACGAAATATTTTATCAAAACAGGATATCTCTGACTATATTC 1952
Db 8201 TTGAAACAGGAGTGATAAAGGAAGAGTCTTTTATCAACAGGATATTTCTATCTACGAAC 8260
Qy 1953 CAGAATATTGCAAGTATAAAAAATTTCTCAGTTTACAAAGAAATGGAACAAATATAATTAACA 2012
Db 8261 CTAAATTTTGTGAGTGTCTCGTTTGTATAGTTTGTGATCAATGAGTAAATTTATGCAAA 8320
Qy 2013 AATCAGAAGTAGTTATTTGCCACGAGGCCCGCTACTTTTATGAAATTCATTTATCCAAAG 2072
Db 8321 AAGCTGATATTATATCACTCACGAGGTCACGCACTTTTATGAGTGCCATTAACAAATG 8380
Qy 2073 GAAAAACCAATTTATTTGTTTCTTAGACAAAAAAGTATGTTGAACATGTAATGATCATC 2132
Db 8381 GCAAAAGCCAAATGTTGTGTCCTAGACAAAGAAAGTTTGGTGAACATGTCATGACCATC 8440
Qy 2133 AAGTAGAGTTTGTAGAAGAAATTTTACAGATAATAATTTTATTTATAGAAAAATATAG 2192
Db 8441 AAGTAGATTTTGCAGAAATGTAGCAAAAGAAATGGAACATCAATTCCTGTTGAAGATG 8500
Qy 2193 ATGATTTGTTGAAAAAATTTATTTGAAGTTTCTAA 2226
Db 8501 TAGATAAATTTGGGGAGACTATTGAAAAATTACAA 8534

RESULT 4
ADSL3200
ID ADSL3200 standard; DNA; 16032 BP.
XX
AC ADSL3200;
XX
XX
DT 16-DEC-2004 (first entry)
XX
DE S. thermophilus NCIM I-2980 polysaccharide synthesis-related operon DNA.
KW lactic acid bacterium; food; meat; cereal; dairy;
KW polysaccharide synthesis; operon; ds.
XX Streptococcus thermophilus.
XX
FH Key Location/Qualifiers
CDS 342..1802
FT /tag= a
FT /product= "Transcriptional regulator eps13A protein"
FT /note= "The corresponding protein sequence is not shown
FT shown within the specification"
CDS 1803..2534
FT /tag= b
FT /product= "Polysaccharide polymerisation/export-related
FT eps13B protein"
FT /note= "The corresponding protein sequence is not shown
FT shown within the specification. The CDS has a prokaryotic
FT Val start."
CDS 2543..3235
FT /tag= c
FT /product= "Polysaccharide polymerisation/export-related
FT eps13C protein"
FT /note= "The corresponding protein sequence is not shown
FT shown within the specification"
CDS 3245..3985
FT /tag= d
```

```
FT FT /product= "Polysaccharide polymerisation/export-related
FT FT eps13D protein"
FT FT /note= "The corresponding protein sequence is not shown
FT FT shown within the specification"
FT CDS 4042..5409
FT FT /*tag= e
FT FT /product= "Undecaprenyl-phosphate glycosyltransferase
FT FT eps13E protein"
FT FT /note= "The corresponding protein sequence is not shown
FT FT shown within the specification. The CDS has a prokaryotic
FT FT Val start."
FT CDS 5611..6195
FT FT /*tag= f
FT FT /product= "Undecaprenyl-phosphate glycosyltransferase
FT FT eps13F protein"
FT FT /note= "The corresponding protein sequence is not shown
FT FT shown within the specification. The CDS has a prokaryotic
FT FT Val start."
FT CDS 6251..6634
FT FT /*tag= g
FT FT /product= "Undecaprenyl-phosphate glycosyltransferase
FT FT eps13G protein"
FT FT /note= "The corresponding protein sequence is not shown
FT FT shown within the specification"
FT CDS 6643..7092
FT FT /*tag= h
FT FT /product= "Beta-1,4-galactosyltransferase eps13H protein"
FT FT /note= "The corresponding protein sequence is not shown
FT FT shown within the specification"
FT CDS 7092..7607
FT FT /*tag= i
FT FT /product= "Beta-1,4-galactosyltransferase eps13I protein"
FT FT /note= "The corresponding protein sequence is not shown
FT FT shown within the specification"
FT CDS 7597..8493
FT FT /*tag= j
FT FT /product= "Rhamnosyltransferase eps13J protein"
FT FT /note= "The corresponding protein sequence is not shown
FT FT shown within the specification. The CDS has a prokaryotic
FT FT Val start."
FT CDS 8763..9797
FT FT /*tag= k
FT FT /product= "Glycosyltransferase eps13K protein"
FT FT /note= "The corresponding protein sequence is not shown
FT FT shown within the specification"
FT CDS 9827..10969
FT FT /*tag= l
FT FT /product= "Repetitive unit polymerase eps13L protein"
FT FT /note= "The corresponding protein sequence is not shown
FT FT shown within the specification"
FT CDS 10984..11793
FT FT /*tag= m
FT FT /product= "Repetitive unit polymerase eps13M protein"
FT FT /note= "The corresponding protein sequence is not shown
FT FT shown within the specification"
FT CDS 11844..12578
FT FT /*tag= n
FT FT /partial
FT FT /product= "Glycosyltransferase eps13N protein"
FT FT /note= "The corresponding protein sequence is not shown
FT FT shown within the specification. No start codon."
FT CDS 12633..13016
FT FT /*tag= o
FT FT /product= "Glycosyltransferase eps13O protein"
FT FT /note= "The corresponding protein sequence is not shown
FT FT shown within the specification"
FT CDS 13049..14482
FT FT /*tag= p
FT FT /product= "Transmembrane transporter eps13P protein"
FT FT /note= "The corresponding protein sequence is not shown
FT FT shown within the specification. The CDS has a prokaryotic
FT FT Val start."
FT CDS complement(14614..15870)
```

```
FT FT /*tag= q
FT FT /product= "Transposase IS1193 protein"
FT FT /note= "The corresponding protein sequence is not shown
FT FT shown within the specification"
FT CDS FR2852604-A1.
FT FT 24-SEP-2004.
FT FT 17-MAR-2003; 2003FR-00003242.
FT FT 17-MAR-2003; 2003FR-00003242.
FT FT (RHOD ) RHODIA CHIM SAS.
FT FT Horvath P, Manoury E, Huppert S, Fremaux C;
FT FT WPI; 2004-671092/66.
FT FT New strain of lactic acid bacteria, useful for preparation of foods and
FT FT pharmaceuticals, especially fermented dairy products, contains genes
FT FT involved in polysaccharide biosynthesis.
FT FT Claim 5; SEQ ID NO 1; 35pp; French.
FT FT The invention relates to a novel strain of a lactic acid bacterium that
FT FT contains at least one of 8 specific nucleic acid sequences all of which
FT FT are defined in the specification. The bacterium of the invention may be
FT FT useful during food preparation, particularly that of beverages and meat,
FT FT cereal, or dairy products e.g. fermented milk, yoghurt, 'matured' cream,
FT FT dessert cream, cheese, soft cheese, cheese spreads, cottage cheese, milk-
FT FT based drinks, dairy product tentates and baby milk, where the products
FT FT are derived from an animal and/or plant. The current sequence is that of
FT FT the Streptococcus thermophilus CNCM I-2980 polysaccharide synthesis-
FT FT related operon DNA of the invention.
FT FT SQ Sequence 16032 BP; 5191 A; 2323 C; 3188 G; 5330 T; 0 U; 0 Other;
FT FT Query Match 12.7%; Score 889; DB 13; Length 16032;
FT FT Best Local Similarity 66.8%; Pred. No. 1.6e-110;
FT FT Matches 1344; Conservative 0; Mismatches 655; Indels 12; Gaps 5;
FT QY 317 ATACCTTTTAACGTAATTAATTAAAGCAGTTTAAAGTAGTCTTCTTATTTTCGACAATCTA 376
FT DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
FT 5592 ATACTTTATCACTGCACTGTGCGTAGATACAAATATCTATTCTTATGACTGCTGAACA 5651
FT QY 377 TCAAAAAGACGATTTCTAATACACGGCTGAAGATGGGAAATATGCAAGTTTATT 436
FT DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
FT 5652 ACAGAAGAATACATTAGTCATTACCACCTGAGAGATTGGGATTAAATGGAGGGCTCTT 5711
FT QY 437 TGAATCACAATAACAATAATTCAAAATAATCTTTGTCATTGCTAGTGTAGTTAGGTACAGAAAT 496
FT DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
FT 5712 TGAGTCAGATCAACTTCTCTATCAAAATATTGGCAGGTATAGTAGTCATCGGAGATGGTGA 5771
FT QY 497 AGATAAAATTAATTTATCATTACCGCTCTATTATTCTGTGAAGAAGCTATAGAGTTTTC 556
FT DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
FT 5772 AGTGGGTTCCAGAGGGAGTCCCAATAATTCCTTTT---GATGATGCGATTGACTTTGC 5828
FT QY 557 AACAGGGAAGTGTGCGACACCGTCTTTTATAAAATCTACCAAGTAGTGTATTTAGAGTAAA 616
FT DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
FT 5829 GACTCATGAAGTTGCGACCATGTGTTTATCAACTTACCGAGTGAACATTACGATCTCAA 5888
FT QY 617 GCATTCGTTTCAGATTTTGAGTTGTTAGTATTTAGTGAAGCTTGATATTAATTCATT 676
FT DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
FT 5889 ACATCTTGTTCGCAATTTTGAAGTCATGGGTATTTGATGTGATGTAGATATTAATTCATT 5948
FT QY 677 CGGTTTTACTCGGTTGAAAACAAAAAATCCAACTGCTAGTGAACCATAGCATTTGTAAC 736
FT DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
FT 5949 TGATTTGAGGCTTTAAAAAATAAAAAATCAACAGTTGAGACCATAGTATGCTGAC 6008
FT QY 737 TTTTTCACAAATTTTATTAAGCCCTAGTCATATCATGATGAAACGACGTTTTGGATATACT 796
FT DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
FT 6009 TTTTAACCTCAATTACTACAAACATAGCCATATCTTTCTAAAGCGCATGTTGGATATCTT 6068
```



QY 133 TTTTATATCTCGTAGCCAGTTGAAATTTAGTATAGAGGTAATCTGATAGAGTTTGAAA 192  
Db 4642 TTTATCTCTCTCGATTTTACAGAGACTTTTGGAGTCGTGGCTATCTTGAAGAGTTTAAAA 46383  
QY 193 AAAACATTAACTATAGTATATATTTGCAATTTTCTTACGGCAGTATCATTTTGTGG 252  
Db 46382 TGGTATTGAAATACAGCTTTTACTATATTTTCAATATCAAGTTCAATATTTTATTTT 46323  
QY 253 AGAATAATTTTCGCACCTTTCAAGACGTTGGTCGGTGTATTTCCACATTAATAAACTTCGTTT 312  
Db 46322 AAAACCTCTTTTACAACGACACGACTTCTCTTTTCTTTTATTTTATGCTATGAATTCGATTT 46263  
QY 313 TGGTATACCTATTTAAACGTAATTAATAGCAGTTTAAGGATAGCTTCTATTTTCGACAA 372  
Db 46262 TATTATCTATTTGAATTTTCAATTTTAAATATTTATCGAAAAATTTCTTTACGCTAAGTTT 46203  
QY 373 TCTATCAAAAAAGACGATCTCAATTAACAACGGCTGAACGATGGGAAATATGCAAGTTT 432  
Db 46202 CACGAGATACCAAGTTGTTTGATAACGAATAAGGATTTCTTTATCAAAAAATGACCTTTA 46143  
QY 433 TATTTGAATCACATAAAACAAATTTCAAAAAATCTTTGTTGCAATTTGGTAGTTTATAGTACAG 492  
Db 46142 GGAATAAT ---ACGACATAATATATCGCTGCTGTATCTTGGATTCTCTGAAAAGG 46086  
QY 493 AAATAGATAAAATTAATTTATCATTAACGCTCTATTTCTGTGGAAGAAGCTATFAGAT 552  
Db 46085 ATTGTTATGATTTGAAAACATAACTCGTTAAGGATAATAAAACAAAGATCTCTTACTTCAG 46026  
QY 553 TTTCAACAAGGAAGTGTGCAACACGCTTTTATAATCTACCAAGTCAGTTTATAGAG 612  
Db 46025 AGTTAACTGCTTAACCTGTTGATCAAGCTTTTATTAACATACCCCAATGAAATTTTGGTA 45966  
QY 613 TAAAGCAATTC-----GTTTCAGATTTTGAGTTGTTAGGTATTCATGTAAGCGTTG 663  
Db 45965 AATACCAATACAGATATTAATGACATTTGAAGCAATGGGAGTATTTGTCATGTTA 45906  
QY 664 ATATTAATTCATTCGGTTTACTCGTTTGA AAAACAAAAATCCAACTGCTAGTGACC 723  
Db 45905 ATGTAGAGGCACTTAGCTTGATAATA TAGGAAAAAGCGAATCCAACTTTTGAAGAT 45846  
QY 724 ATAGCATTTGAACCTTTTCCAAATTTTATAGCCCTAGTCATATCATGATGAACGAC 783  
Db 45845 ATAGTGTATTACATATTTCTATGAATTTCTATAAATATAGTCACCTTATAGCAAAACGAT 45786  
QY 784 TTTTGGATATACTCGAGCGGTAGTCGGTTAAATTTATTTGGTATAGTTCTATTTTGT 843  
Db 45785 TTTTGGATATCAGGGTCTATATFAGTTTGTCTCAATGTCGCAATGTCGCAATTTTTC 45726  
QY 844 TAGTTCCAAATTAATCGTAGAGATGGTGACCGGCTATTTTGTCTCAGAAAACGAGTTGGAC 903  
Db 45725 TAGTTCCAAATCAGAAAAGATGGTGACCGGCTATCTTTTCTCAAAATAGAGTAGGTC 45666  
QY 904 AGAATGGACGATTTTACATTTCAAGTTTCGATTCGATGATGTTGATGCTGAGGAGC 963  
Db 45665 GTAATGGTAGGATTTTATAGTTCTATAAATTCAGATCAATGCGAGTAGATGCAACAAA 45606  
QY 964 GCAAAAAGACTGTCTAGCCAAACACAGATGCAAGGTTGGGTATGTTTAAATGGGAA 1023  
Db 45605 TTAAGAAAGATTATATAGTTCAATCAATCAATGACAGGG-----CTAATGTTTATAGTTAGA 45551  
QY 1024 AAACGATCCTAGAAATTTACTCAATTTGGACATTTTATAGCCAAAAACAAAGTTTATAGCAGT 1083  
Db 45550 CGATGATCCTAGAAATTAATAAATAGGAAATTTATTCG-AAAAACAGCATAGATGAGT 45492  
QY 1084 TACCAAGTTTATTAATGTTTAAATTTGGCGATGATGAGTCTAGTTGGGTACAGTCCACCTA 1143  
Db 45491 TGCCTCAATTTCTATAATGTTTAAAGGGTGTATGAGTTTATAGTGAACAACCGCTCCCA 45432  
QY 1144 CAGTTGATGAATTTGAAAATAATATCTCTGCTCAAAAGAGAGGATTTAGTTTAAACCGAG 1203  
Db 45431 CAGTTGATGAATGAAAAGTATTAATTAACGCAAGACGAGCTTAGTTTATAGCCAG 45372  
QY 1204 GGATTTACAGGTCCTCGGCAGGGTTAGTGTGCTAGTAAATATACAGACTTCGACGAGCTAG 1263

Db 45371 GAATCACTGGTTTGTGGCAAAATATCTGCTAGAAAATAATATTTACTCATTTTGTGATAATCG 45312  
QY 1264 TTCGGTTTGGACTTAGCATACATTTGATTAATTCGACTATCTGCTCAGATATATAAAATTTTAT 1323  
Db 45311 TAAAGTTAGATGTTCAATATATCAATGNAUGTCTATTTTGGTCAGATATTAAGATTATTC 45252  
QY 1324 TAAAGACAGTGAAGTTGTATTTGTGTAGAGAGGGAAGTAAGTAAAGTAT----- 1373  
Db 45251 TCCTAAACACTAAAGGTAGTTTCTACGCGGACAGGAGCTAAGTAAAGGTAAAGGTTTGAAG 45192  
QY 1374 -----ATGAAAGCTTTTGTGTCGGTCTTTCAGGGGACATTTGACTCACTTGTATTT 1426  
Db 45191 GAATATATAGAAAATTTGTCTGGTTGGTTCAGGTGGTGGTCACTAGCACACTTTGAACCT 45132  
QY 1427 GTTAAAAACCGTTTGGGAAGGAAGAACGTTTTTGGGTAAACATTTGATAAAGAGAGATGC 1486  
Db 45131 TTTGAAACCCATTTGGGAAAGAAAGATAGGTTTTGGGTAACTTTGATAAAGAGATGC 45072  
QY 1487 AAGAGCTTTTGAAGAAATGAAAAATGTATCCATGTTTACCTTCCAAACAAATCGCAATCT 1546  
Db 45071 TAGGATATTCTAAGAGAGAGAGATTGTATATCAATGCTTCTTCCAAACAAACCGTAAATG 45012  
QY 1547 CATTAATTTAGTGAATAATCTTTCTTAGCTTTTCAAAATTTTACGTGATGAGAACCCAGA 1606  
Db 45011 CAAAACCTTGGTAAAAAATACTATTTCTAGCTTTTAAAGTCTTGA AAAAGAAAGAACCCAGA 44952  
QY 1607 TGTATTATTATCTCTGTCGGCGGCTGTCGTCCCTTCTTTTACATCGGAAACATATT 1666  
Db 44951 TGTATCATATCATCTGTCGGCTGTAGCAGTACCATTCTTTTATATTTGTTAGTATTT 44892  
QY 1667 TGGAGCAAGACGATTTATATTTGAAGTATTTGATCGAGTTAATAAATCTACATTAACCTGG 1726  
Db 44891 TGGTTGTAAGACCGTTTACATAGAGGTTTTCGACAGGATGATATAACCACTTTCACAGG 44832  
QY 1727 AAAACTAGTTTATCCCGTAAACAGATATTTTATTTGTTTCAGTGGGAGAAATGAAGAGGT 1786  
Db 44831 AAAATTAGTGTATCTCTGTAAACAGATAAATTTATTTGTTTCAGTGGGAGAAATGAAAAAGT 44772  
QY 1787 ATATCTTAATCTATTAACCTTGGGAGTATTTTATGATTTTGTAACTAGTAGAACT 1846  
Db 44771 TTATCTTAAGCAATTAATTTAGGAGGATTTTATGATTTTGTACAGTGGGGACA 44712  
QY 1847 CATGAACAAACAGTTTAACTCGATTGATAAAGAGATTCATTTTATGAAAAAATGGAAGT 1906  
Db 44711 CATGAACAGCAGTTTCAACCGTCTTATTAAAGAGTTGATAGATTAAAGGACAGGTGCT 44652  
QY 1907 ATAAACGACGAAATTTATTTTCAACAGGATTTCTGACTATATTTCCAGAAATTTGCAAG 1966  
Db 44651 ATTGATCAAGAAGTGTTCATTTCAACCGGTTTACTCAGACTTTGAACTTCAGAAATTTGTCAG 44592  
QY 1967 TATAAAAAATTTCTCAGTTTACAAAAGAAATGCAACAAATATATAAATAATCAGAACTAGTT 2026  
Db 44591 TGGTCAAAATTTCTCTCATATGATGATGAACTCTTACATGAAGAAGCTGAGATTGTT 44532  
QY 2027 ATTTGCCACGAGGCGCCGCTACTTTTATGAATTCATTTATCCAAAGGAAAAAAACAAATTA 2086  
Db 44531 ATCACAATGCGCGCCAGCGCGTTTATGATGCAAGTTTCTTAAAGGAAAAAATACTATT 44472  
QY 2087 TGTGTTTCTAGACAAAAAAGTATGGTGAAACATGTAAATGATTCATCAAGTATAGATTGTA 2146  
Db 44471 GTGGTCTCTAGACAAGAACAGTTTGGAGAGCATGTGAATAATCATCAGGTGGACTTCTGT 44412  
QY 2147 AGAAGATTTTACAAGATAATAATTTATTTATAGAAAATATAGATGATTTGTTTGA 2206  
Db 44411 AATAAGGTAAAAACAATGTATTAATTTTATGATTCGTTGTAGATATTTGAAGGTTTACAAAT 44352  
QY 2207 AAAATTTATGAGTTTCTTAAGCAAACTAACTTTTACATCAAAATAATAATTTT---TTTTGT 2263  
Db 44351 GTAGTCTATAGGGAATGATGATCGTCGGTTTTTTAGAACTAATAGTAGTAATTTTATTT 44292  
QY 2264 GAAAGATTAAAAACAAATAGTTTGA AAAATTTTAAATGAGGATCAAGAAAAATGAATAATAAAAA 2323

Db 44291 GAAGAAATTAAAGTAATATTAAAGGAGTTGGCGATGAAATCAATAAAAACTCTTTATT 44232  
Qy 2324 AGATGCATATTTGATAATGGCTTATCATATTTTCTCAGATTTTACTCGAGAGGGATAC 2383  
Db 44231 TTATATTGCAATATTTTGTAGTAAATTTTAAATCACTGGGTTTAGCGGAGGAAACTC 44172  
Qy 2384 AGATATTATCATCTTCTCTCAGGAGATGACACCATTTAGTTCCTTCAGAATCCTGTAT 2443  
Db 44171 AGCTTACAAATAGTG-----ATGTTAGTTCGAA 44143  
Qy 2444 AATTATTAAATTTCTCAGATTTATATGTTGAAATTTTACAAAGATGAGCAAAAATAT 2503  
Db 44142 TTTTACTGTGGGAATAAAATTTTTTATAGATAGCTTTTATTTTGAAGAAGAAACTCG 44083  
Qy 2504 AAAGAAATAGGATATATGAAGAGTTAAATGTTTACAGATTTATTTCCATAATATACGAA 2563  
Db 44082 TGATCATCTTTTATTTATTTATTCGCGACCAATTTTGAATTTTATCT----- 44038  
Qy 2564 AAAACTATTGATAATGTACTGTTTAGAATTTTATTAAGAAATGATCGAGCTTTTGAATAC 2623  
Db 44037 ----TTGTTCTAAGGTTACTTTTATATTAATTAATTTTCTTAGCATTAAGGAC 43982  
Qy 2624 TATTACAAAGATTTGTTTATTTGATAGATAAAAAAATGCTGCTAAGAAATAGATTTG 2683  
Db 43981 ATCTCTCAAAAAGCTTTCTCTATAATATAGATCGCGTATTTTGGGAGTTCTATTA 43922  
Qy 2684 GTTCTAATTTGGGTTTCGCTCCACAGATTTTGGGCAATTCCTTTATCAAAATGAAACG 2743  
Db 43921 AATCAAAATTTTGTGAATTAGATTTAATAGAAATTAAGTATGTCAATTTTATAGGGAT 43862  
Qy 2744 AAACAGCTTATTTATTAAGTAATCTAAATCTCCAGATGAACTATTTATCAGACAAATTA 2803  
Db 43861 GGACAAATTTATCGAGAAGTGA-CCTAGGTTTGGTCATCCTAACTTTATTCATAATTT 43803  
Qy 2804 TAGAAAAATATGAATTTTCAATAGATTAATCTAAATATGAA-----ATTAAAGATATATA 2860  
Db 43802 TTTTGCCTAACTATTTCTGTATATTTGTACTCAATTAATAACGACTAAAGCCTGTTGT 43743  
Qy 2861 AGTGAAAAAATCAACATCTTCTCCTATGTTCTTTTACAGATGATCTTATGATGAATGC 2920  
Db 43742 GATGGTTTATTTTAAATTAATTAATTTATTTGTCACCAATATCTTTTCAAGGACAGG 43683  
Qy 2921 TAAATGCAAGAAATTTAGGTTTTTATTTGCTAGAAAGTTAAAAATAGAAATAAATCTA 2980  
Db 43682 GTATTATATCGTAATTTTATTTATTTGTTACTCAATTTATGTGCAAGAAATAGCTTAATAA 43623  
Qy 2981 AA----TTTAAAGAAATTTACTAAAAAATAAAATAGTTGATTTTGTGAGAGTAATGTA 3036  
Db 43622 AAGAGTATTTATGAAATTTAGCACCTATGTACAAATTTTTTTATAGTATTTACCTTTTT 43563  
Qy 3037 TGTTTAAATTTTAAATATGACCCGGGAATATTTTATTTTAAAGTACTTCTGGTTGATTA 3096  
Db 43562 GAGTCTCAAAATTTTTTAAATCAAAATTTGTTCAAAAATTTAGATGTTCTTTTAAACAGG 43503  
Qy 3097 TTTTATTTCCAGACAAAGTATGTTATTTTATTTTAAATTTTATGAAATTAATTTATTTTC 3156  
Db 43502 TAGATTACATATGCTCATTTTCAACCTTTGATAGTGGTTTAACTCCTTTTGGAAATAGTTT 43443  
Qy 3157 ATATAAAATTTTGAATACTAAGCTAATATTTTAAAAAATGAAATTTTATTTTATTTAT 3216  
Db 43442 TAAGGAAACAAGTGTCTTATTTGATATAGCTACTCTATGTTATTTGATATGATGGTGT 43383  
Qy 3217 GGTCTATATTATGTTTGTTCAGTGTGACAAAGTATGTTTGTGAAATAAATTTTGA 3276  
Db 43382 AGTACTTACCATGTTTGTATGATAATCTATATATCTATATAGTAAAAA-----TAA 43330  
Qy 3277 GATTATTTGAGATTTTACTGCTCCCAATAATTTGGAATTTTGAATATGATTTATTAAT 3336  
Db 43329 TCATAATTTGAATCTCAACTACTCTCTATTTATTAATGTCTATATATATTTTACTGAAAGTT 43270  
Qy 3337 TGTATTCAATTTATAAATATGATTTATAAAAAATTTAAAAAATAGTATCTCTTTTATGTTTTT 3396  
Db 43269 TTTATCCAGTGTGTAATGAATATTAGTTGGCTAGTTTTTGGTAAAAATATTTTGTGATG 43210

Qy 3397 TACGTTTTTATAGGTATATCTGCATTTGTATATTATTCAAAATGGAAAGATATTGCTATTTT 3456  
Db 43209 GTATCGAACCTATAAAAAAGGAATTTACTATTGTGAATAAT--ATATGACATATTTGCTC 43152  
Qy 3457 TAGACAGACACCTTATAGGACTAGACTATCTTATAACAGCGCTCAAAACAAGGTTGGTTG 3516  
Db 43151 TGATATGGCAGGAGGTAAAGAAAGAAATGATACCTTAAAGTTATACATATTGTTGGTTT 43092  
Qy 3517 GCTTTATGAACACTACCTTACCGTTAAATACCACTACAAATATATAGTTTCAATTTCCGTTAATCT 3576  
Db 43091 GGAGGAAATCCCTTACCAGATAATTTAAAGAAATATATAAAAACTTGGAGAGAAACAATGT 43032  
Qy 3577 TTGCACCTATAAAAATAAATGCAACAAATTTTTTTCTTGTGCTTGTCTTTTATACCGA 3636  
Db 43031 CCGGATTTATGAAATTTATGAATGGAATGAGCATAAATTTATGATGT----- 42988  
Qy 3637 TCTATTTTAAGTGGATCGAGAAATTTGGTAGTTTATCGCTAGCAATATTAATTTATATGCTTGT 3696  
Db 42987 --TAGTAAAAATGTTTTTATGAGAGAGCAATATACTAAGAAGAAATTTTGGCTTATGTTTCT 42930  
Qy 3697 TATGGAGATATATAGG--TGGAAAAATTTGCTTTGGATAAAAAAGCTAAATAGTAATATTTGTA 3755  
Db 42929 GACTATGCAAGATTTGGATATTTATTTATACITTTATGGGGGTTCTATCTAGATACCTGATGTG 42870  
Qy 3756 ATACTACTTATTTTAAATACTGAATGCTTTTACATGAAATTTTGGCTGTTGTTTAAAT 3815  
Db 42869 GAGCTTTTAAAAAGTTTGTAGTCTCTTTGAGGATTTATGAGTGTCTTTTACGAAGGAGATT 42810  
Qy 3816 TCTAGAGAAATCAAGTACGAAGCTAGATTTTATTTATCAAGGAAGTATTGTATAAAGTA 3875  
Db 42809 AGTTGATGTGAATACAGGATTAATTAATTTGGCGCTGTTAAAGGACATCACTTTTAAAA 42750  
Qy 3876 TTAGAAAAACAATATTTTATTTGGATATGGAATATCCGAATATTTCAGTTACGGGAACCTGG 3935  
Db 42749 TCAATATATGCTATATATACAAAAAGTGATTTAACTTCTCTTAATAAGACATGTGTAGAG 42690  
Qy 3936 CTGGAAAGTCAATTCAGGCTATATATCAATTTTTTTTAAATCAGGAATAGTTGGGTTGAT 3995  
Db 42689 GTTACAACTAATTTTATGATAAACAAGGCTTTAAAGAAAT--AAGAAATATTATTCAAAAGA 42632  
Qy 3996 TTACTGATGTTTCTTTTTTTTTTATGTTATATAAAAAAGTTATCGAGTTAATGGGAAACA 4055  
Db 42631 TTGATGATAAACAATATATCCGAGAAATTTTAAATCCAAAGAAATTTATTAACAGGTA 42572  
Qy 4056 GCCTATTTTATTTTATCATCATTTAGCCATATTTTTCATATATGAAACAATAGATCCGAT 4115  
Db 42571 AGGTTGATTTCTGACTAGTGTTTACCTATTTCTATACATCATTTACGAAGGAGTTGGAAAA 42512  
Qy 4116 ATTATTATATATAGTACTATTTCTTTTCTTCAATAGGTTATTTGGAAATAATATAAATTTTAAA 4175  
Db 42511 GTTCTTCATTTATTTTCAAGATCTCTAAAGATTTAGAGTAAGGCTCATTAATTTGATTTTTAT 42452  
Qy 4176 AAGGATATGAGACAAAAAATGAATGATTTTAAATTTTCAGTTATTGTACCAATTTTATATGT 4235  
Db 42451 TTGGATATGTTACTTATAGAAATGCTTCTAAGGTTTCTTAAAGTTTAAAGAAATAGTTATAT 42392  
Qy 4236 CCAAGATTTATCTTGATAAATGTTATTAACAGTATTTTAAACCAACAATATATACTAATTTAGA 4295  
Db 42391 TGTAAATAGAAGCGAAATGTTATATATTTTACTCTTACATTTTAAACAGGCTACCGA 42332  
Qy 4296 GGTATTTCTCGTAAATGATGGAAGTACTGATGATTTCTGAGAAATTTGCTTTAAACTATAT 4355  
Db 42331 TTATCCTATC-TGTATGATTTCTCTATGTAATCAGACAAATAAAAAATTTTATATGTTGAT 42273  
Qy 4356 GAAGACGATGGAAGAATTAATATATTAAGAATAATTAATGGCGGTCTAGCAGATGCTCG 4415  
Db 42272 TGTGATGATGGTCTCGAGGACAGTACTAAGGAAATTTGTAAGTAATTTATATAAAGAGAA 42213  
Qy 4416 AAATTTCCGACTAGAAACATGCAACAGGTAATAATATTTGCTTTTGTCCGATTTCTGATGACTA 4475  
Db 42212 TAAAGTTAGTATTGTATATCTGTA--TAAACGTAATGGCGGTAAAGCAATTCAGCCTATAA 42156





Db 40038 GCTATTTTACTGTAGCACATAATAAATATATATAAGAGAGAAATTAATTTCTACGCTGTC 39979  
Qy 6687 TTGAAAAGGTAGATTCATGAAGATGAATCTTCACTTATCCCTTGCTCTATGAGTTA 6746  
Db 39978 TATCCAGTAGGGAAGTTACACGAAGATGAATTTTAAACCTATATAATTTTGAAGCC 39919  
Qy 6747 GAAAAAGTTGCAATAGTTAAGGAGTGTCTGTACTATTATGTTGACCGAGAAATAGTATC 6806  
Db 39918 AAAAAATATAATTTCTTTAGATATAATACATTATGCAATATGTAATAGAGAAAATAGCATA 39859  
Qy 6807 ACAACTTCTAGCATGACTGACCATCGCTTCCATTCGCTACTGGAATTTCA 6856  
Db 39858 ATGACTGTTCTCTACATAATTTAAAGGCTACATGCTGTAGAAGCACTTAA 39809

## RESULT 6

ID ABK90550  
AC ABK90550 standard; DNA; 6865 BP.

AC ABK90550;

XX 15-NOV-2002 (first entry)

DE Betal,3-galactose transferase DNA #2.

XX Betal,3-galactose transferase; gene; ds; galactose; N-acetylglucosamine;  
KW betal,3-bond; galactose-containing glucide; uridine-5'-diphosphate;  
KW infection.

XX Streptococcus agalactiae.

PH Location/Qualifiers

FT CDS 617..1792

FT /\*tag= a

FT /product= "Betal,3-galactose transferase #2"

FT 1816..2262

FT /\*tag= b

FT /product= "Betal,3-galactose transferase #3"

FT /note= "No stop codon given"

FT 2265..2747

FT /\*tag= c

FT /product= "Betal,3-galactose transferase #4"

FT 2843..3979

FT /\*tag= d

FT /product= "Betal,3-galactose transferase #5"

FT /partial

FT /note= "No stop codon given"

FT 3982..4956

FT /\*tag= e

FT /product= "Betal,3-galactose transferase #6"

FT 5009..5950

FT /\*tag= f

FT /product= "Betal,3-galactose transferase #7"

PN JP2002199885-A.

XX 16-JUL-2002.

XX 05-JAN-2001; 2001JP-00000392.

XX 05-JAN-2001; 2001JP-00000392.

XX (KYOW ) KYOWA HAKKO KOGYO KK.

XX WPI; 2002-612563/66.

XX P-PSDB; ABG31703, ABG31704, ABG31705, ABG31706, ABG31707, ABG31708.

XX Betal,3-galactose transferase and a DNA encoding the enzyme, a protein  
FT having betal,3-galactose transferase activity, a transformant,  
PT preparation of a protein, and preparation of a galactose-containing  
PT glucide.

PS Example 1; Page 15-22; 24pp; Japanese.

XX The invention relates to a protein having betal,3-galactose transferase  
CC activity, derived from a microbe having an activity of transferring  
CC galactose to N-acetylglucosamine by betal,3-bond. The protein can be used  
CC in a method for the preparation of a protein having betal,3-galactose  
CC transferase activity in which the above transformant is cultured in a  
CC medium to form and accumulate a protein having betal,3-galactose  
CC transferase activity, and in a method for the preparation of a galactose-  
CC containing glucide in which a culture liquid of the above transformant or  
CC a treated product of the culture liquid is used as an enzyme source, the  
CC enzyme source uridine-5'-diphosphate galactose and a receptor glucide are  
CC made to be present in an aqueous medium, and a galactose-containing  
CC glucide is formed and accumulated in the aqueous medium. The galactose-  
CC containing glucide is used as a candidate for infection preventors. This  
CC sequence represents DNA encoding betal,3-galactose transferase proteins  
CC of the invention

SQ Sequence 6865 BP; 2453 A; 847 C; 1185 G; 2380 T; 0 U; 0 Other;

Query Match 10.4%; Score 727.8; DB 6; Length 6865;  
Best Local Similarity 59.2%; Pred. No. 6.9e-89;  
Matches 1389; Conservative 0; Mismatches 917; Indels 41; Gaps 7;

Qy 16 TGGCATTATTGATATGATAGACAGTTGCAATTTCTGCAATCTTAAACAAGTCATATACCAA 75  
Db 438 TAGCGATGATTCAACACGGTTGGTATTATTTTCTGCAAGTTTGACATTAACATTAATTA 497  
Qy 76 ATGCTGATTTAAATCG---TTCTGGAATTTTATCATATGATGATGTTTCATTATTTTGCAT 132  
Db 498 CTCCTCACTTTAAAGCAATAAGATTATTTGTTGTTCTATTGATACATATATTTGTCT 557  
Qy 133 TTTTATATCTCGTATGCCAGTTGAAATTTGAGTATAGAGTAATCTGATAGAGTTGAAA 192  
Db 558 TTTATCTTCTGATTTTACAGAGACTTTTGGAGTCGTGGCTATCTTGAAGAGTTTAAAA 617  
Qy 193 AAACATTTAACTAGTATATAATTTTGCATTTTCTTACGGCAGTATCATTTTGTGTTG 252  
Db 618 TGGTATTTGAAATACAGCTTTTACTATATTTTTCATATCAAGTTTCATTATTTTATTTCTA 677  
Qy 253 AGAATAATTTGCGACCTTCAAGACGTTGGTGGCGGTGATTTTTCACATTAATAAATCTGTTT 312  
Db 678 AAACTCTTTTACACGACGACTTTCCTTTTACTTTTATTTGCTATGAATTCGATTT 737  
Qy 313 TGGTATACCTATTAAACGTAATTTTAAAGCAGTTTAAAGATAGCTTTCTATTTTCGACAA 372  
Db 738 TATTATATCTATTGAAATTCATTTTAAATAATATATCGAATAATATCTTACGCTAAGTTT 797  
Qy 373 TCTATCAAAAAGACGATTTCTAATTACACGCTGAACGATGGGAAATATGCAAGTTT 432  
Db 798 CACGAGATACCAAGTTGTTTGTATACGAATTAAGGATTTCTTATCAAAAATGACCTTTA 857  
Qy 433 TATTTGAATCACATAAAACAAATTTCAAAAAATCTTGTTCGATTTGGTAGTTTATAGGTACAG 492  
Db 858 GGAATAAAT---ACGACCAATAATATATCGCTGCTGTATCTTGGATCTCTCGAAAGG 914  
Qy 493 AAATAGATAAAATTAATTTATCATTTACCGCTCTATTTATTTCTGTGGGAAGAGCTATAGAGT 552  
Db 915 ATTGTTATGATTTGAAAACATAAATCGTTAAGGATAATAAAACAAAGATGCTCTTACTTTCAG 974  
Qy 553 TTTTCAACAGGGAAGTGGTCGACCAAGTCTTTTAAATCTTACCAAGTGAAGTTTGTAGACG 612  
Db 975 AGTTAACTGCTTAACTGTTTGAATCAAGCTTTTATTAACATACCAATTTGATTTTCGGTA 1034  
Qy 613 TAAAGCAATTC-----GTTTCAGATTTTTCAGTTTGTAGGTATTTGATGAAGCGTTG 663  
Db 1035 AATACCAATACAGATATTAATTAAGACATTCAGCAATGGGAGTGAATTTGTCAATCTTA 1094  
Qy 664 ATATTAACTCAATTCGGTTTTTACCTGGGTTGAAAAACAAAAATCCAACTCGTAGGTGACC 723  
Db 1095 ATGTAGAGGCACCTAGCTTTTGATAATATAGGAGAAAAAGCGAACTCCAACTTTTGAAGAT 1154  
Qy 724 ATAGCATTTGTAACCTTTTTCACAAATTTTATTAAGCCTAGTCATATCATGATGAACGAC 783

```
Db 1155 ATAGTGTTATTACATATTTCTATGAATAATCTATAAATATAGTCACTTATAGCAAAACGAT 1214
Qy 784 TTTTGGATATACCTCGAGCGGTAGTCGGGTTAAATTTATTTGGTATAGTTCTATTGTTGT 843
Db 1215 TTTTGGATATCATCGGTCTATTATAGTTTGGCTCATATGTCGCAATTTGCGCAATTTTC 1274
Qy 844 TAGTTCGAATTTATCGTAGAGATGTCGACCGGCTATTTTGGCTCAGAAAAGAGTTGGAC 903
Db 1275 TAGTTCGCAAAATCAGAAAAGATGTCGACCGGCTATCTTTTCTCAAAATAGAGTAGGTC 1334
Qy 904 AGAATGGACGCATATTTACATTTCAAGTTTCGATCGCATGATGTTGTCATGCTCAGGAGC 963
Db 1335 GTAATGGTAGGATTTTATAGTTCTATAAATTCAGATCAATCGAGTAGATGCGAGAACAA 1394
Qy 964 GCAAAAAAGACTGTCTAGCCAAAACACAGATCGAAGGGTGGGTATGTTTTAAATGGGAA 1023
Db 1395 TTAGAAGATTTATTAGTTCACAATCA-----AATGACGGGGCTAATGTTTAAAGTTAGA 1449
Qy 1024 AAACGATCCTAGAAATTTACTCCAATTTGGACATTTTCATACGCAAAAACAAGTTTAGACGAGT 1083
Db 1450 CGATGATCCTAGAAATTTACTAAAATAGGAAAATTTTATTCG-AAAAACAAGCATAGATGAGT 1508
Qy 1084 TACCACAGTTTATATAGTTTAAATGCGGATATGAGTCTAGTTGGTACACGTCCACCTA 1143
Db 1509 TGCCTCAATTCATAATGTTTTTAAAGGTGATATGAGTTTATAGGAAACACGCGCCCTCCA 1568
Qy 1144 CAGTTGATGAATTTGAAAAATATATCTCTCGTCTCAAAAGAGACGATGAGTTTAAACCCAG 1203
Db 1569 CAGTTGATGAATTAAGAAAAGTATATATCAACGAGACGACGCGCTTATGTTTTAAGCCAG 1628
Qy 1204 GGATTACAGGTCCTCGCAGGTTAGTGGTCTGTTAGTAAATATACAGACTTTCAGCACGCTAG 1263
Db 1629 GAATCACTGGTTTGTGGCAATATCTGCTAGAAATAATATTTACTGATTTTGTATGAATCG 1688
Qy 1264 TTCGGTTGGACCTTAGCATACATTTGATTAATTTGGAATATCTGCTCAGATATTAATTTTAT 1323
Db 1689 TAAAGTTAGATGTTCAATATATCAATGAATGCTCTATTTGGTTCAGATATTAAGATTATTC 1748
Qy 1324 TAAAGACAGTGAAGTTCTGTTGTTGAGAGGGAAGTAAAGTAAAGTA-----1372
Db 1749 TCCTAACGCTAAAGTAGTCTTTTACTCGGGACAGGAGCTAAGTAAAGGTTGAAAG 1808
Qy 1373 -----TATGAAAGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1426
Db 1809 GAATATATGAATAATTTCTGTTGTTGTTCAAGTGGTGTCACTAGCACACTTGAACCT 1868
Qy 1427 GTTAAACCGTTTGGAAAGGAAGAAACGTTTGGGTAAACATTTGATAAAGAGGATGC 1486
Db 1869 TTTGAAACCCATTTGGGAAAAGAAAGATAGGTTTGGGTAACTTTTGATAAAGAGATGC 1928
Qy 1487 AAGAGTCTTTGAGATGAATAATGATCCATGTTACTTTCACAAATCGCAATCT 1546
Db 1929 TAGGAGTATCTTAAGAGAGAGATGTTGATATCATTTGCTTCTTCCACAAACCGTAATGT 1988
Qy 1547 CATTAATTTAGTGAATAATCTTCTTAGCTTTCAAAATTTTACGTGATGAGAACCCAGA 1606
Db 1989 CAATAACTTTGTAATAAATACTATTTCTAGCTTTTAAAGTCTTGAATAAAGAAAGCCAGA 2048
Qy 1607 TGTTATATTTTATCTGTTGGCGCGTTGCTGTCCCTTTCTTTTATCATCGGAAACATTT 1666
Db 2049 TGTTATCATATCATCTGTTGGCGCTGTAGCAGTACCATTTCTTTTATATTTGGTAGTTATT 2108
Qy 1667 TGGACAAAGAGATTTATATTAAGATTTTGGATGATGATTAATAATCTACATTAACCTGG 1726
Db 2109 TGGCTGTAAAGCCGTTTATATAGAGGTTTTCGACAGGATAGATAACCAACTTTGACAGG 2168
Qy 1727 AAAAGTACTTTATCCGTAACAGATATTTTATTTGTTTCAGTGGGAAGAAATCAGAAGGT 1786
Db 2169 AAAATAGTGTATCTCTGAACAGATAAATTTATTTGTTTCAGTGGGAAGAAATGAATAAGT 2228
Qy 1787 ATATCCTAAATCTATTAACTTTGGGAGGTATTTTAAATGATTTTGTAAACAGTAGGAACT 1846
```

```
Db 2229 TTATCTTAAGCAANTTAATTTTAGAGGAAATTTTAAATGATTTTTTGTACAGTAGGGACA 2289
Qy 1847 CATGAACAAACAGTTTAAATTCGATTCGATTAAGAGAGATTTGATTTTATCAAAAAAATCGGAAGT 1906
Db 2289 CATGAACAGCAGTTCAACCGTCTTATTAAAGAGTTGATAGATTAAAGAGGACAGGTGCT 2348
Qy 1907 ATAACCGACGAAATATTATTATTTCAACAGGATTTCTGACTATATTTCCAGATATTTGCAAG 1966
Db 2349 ATTGATCAAGAAGTGTTCATTTCAACCGGTTTACTCAGACITTTGAACTTCAGAAATTTGTCAG 2408
Qy 1967 TATAAAAAATTTCTCAGTTTACAAAGAAATGCAACAAATATATAACAAATCAGAAGTAGTT 2026
Db 2409 TGGTCAAAATTTCTCATATGATGATGATGAATCTTTACATGAAGAAGAGCTGAGATTGTT 2468
Qy 2027 ATTTGCCACGAGCCCGCTACTTTTATGAATTCATTTATCCAAAGGAAAAAACAATTA 2086
Db 2469 ATCACACAGCGCGTCCAGCAACGTTTATGAATGCAGTTTCTAAAGGAAAAAACAATTT 2528
Qy 2087 TTGTTTCTTAGCAAAAAAAGATGTTGGTGAACATGTAATGATCATCAAGTAGAGTTTGT 2146
Db 2529 GTGGTTCTTAGCAAAAAAAGATGTTGGAGAGCATGTGAATAATCATCAGGTGGAATTTTGT 2588
Qy 2147 AGAAGATTTTACAGATTAATAATTTTATTTATAGAAATATAGATGATTTGTTTCA 2206
Db 2589 AAGAGTTATTTCTTTGAAATATGAGTTAGATTATATTTTGAATATCAGTGAATTAGAGAT 2648
Qy 2207 AAAATTTATGAAGTTTCTTAAGCAAAAC---TAACTTTTACATCAAAATTAATTTTTTTTGT 2263
Db 2649 ATTATTAAGGAAAAAATAATATCTACTAGTAAAGTAAATATCACAAACAATGATTTTGT 2708
Qy 2264 GAAAGATTAAACAAATAGTTGAAAAATTTTAAATGAGGATCAAGAAAAATGAATAATAAAAA 2323
Db 2709 TCCTCTTTCAAAAAATGAACCTTTCTAAACTATTTGAATAAATAATATTTTGTGGAGAAAA 2768
Qy 2324 AGATGCA 2330
Db 2769 AATTGAA 2775

RESULT 7
ID ACA64723 standard; DNA; 17276 BP.
XX ACA64723;
AC ACA64723;
DT 18-JUN-2003 (first entry)
XX Streptococcus capsular polysaccharide gene.
DE Superantigen; ds; gene; SAg; staphylococcal enterotoxin; tumour; cancer;
KW apoptosis; gene therapy; mammalian cell receptor; cytostatic;
KW tumour associated lipid; anergy; T cell; antigen presenting cell; APC;
KW tumouricidal immunocyte; antitumour.
XX Streptococcus sp.
OS US2002177551-A1.
PN 28-NOV-2002.
PD 30-MAY-2001; 2001US-00870759.
XX 31-MAY-2000; 2000US-0208128P.
PR (TERM/) TERMAN D S.
XX Terman DS;
PI WPI; 2003-361759/34.
DR P-PSDB; ABU79115.
XX A mammalian cell receptor, useful in the treatment of cancer by binding
PT to tumor associated lipids where the binding induces anergy or apoptosis
```



```
Db 6016 TAAAGTTAGATGTTCAATATATATCAATGAATGGTCTATTTTGGTCAGATATTAAGATTATTC 6075
Qy 1324 TAAAGACAGTGAAGTCTGATTTGTTGAGAGGGGAAGTAAGTAAAGTA----- 1372
Db 6076 TCCTAACACTAAAGGTAGTTTACTCGGACAGAGGACTAAGTAAAGGTAAAGTTGAAAG 6135
Qy 1373 -----TATGAAGTTGTTGGTTCGTTCTTCAGGGGACATTTGACTCACTTGTATTT 1426
Db 6136 GAATATAATGAATAATTTCTGTTGGTTCAAGTGGTGCATCTAGCACACTTTGAACCT 6195
Qy 1427 GTTAAACCGTTTTCGAAGGAAGAAGACGTTTTCGTTGTAACATTTGATAAAGAGAGTGC 1486
Db 6196 TTTGAACCCATTTGGGAAGAAGAGATAGTTTTCGTTGTAACATTTGATAAAGAGATGC 6255
Qy 1487 AAGAAGTCTTTTGAAGAATGAATAATGATATCCATGTTTACTTTCACAAATAATTCGAATCT 1546
Db 6256 TAGGAGTATCTAAGAGAAGAGATGTTATATCATTTGCTTCTTTCACAAACCGTAATGT 6315
Qy 1547 CATTAATTTAGTGAATAATCTTCTTAGCTTTCAAAATTTTACGTGATGAGAACCAGA 1606
Db 6316 CAAAACCTTGGTAAATAATCTATTTCTAGCTTTTAAAGGTCCTTGAAGAAAGAACACAGA 6375
Qy 1607 TGTATTATTTATCATCTGCTGGCGGCTGCTGCTCCCTTCTTTTACATCGGAACACTATT 1666
Db 6376 TGTATCATATCATCTGCTGGCGGCTGCTAGCAGTACCATCTTTTATATTTGGTAAGTTATT 6435
Qy 1667 TGGAGCAAGACGATTTATTTAAGTATTTGATCGAGTTAATAATCTACATTAACCTGG 1726
Db 6436 TGGTTGTAAGACCGTTTATATAGAGTTTTCGACAGGATAGATAAACCACTTTCACAGG 6495
Qy 1727 AAAACTAGTTATCCCGTAACAGATATTTTATTTGTTTCAGTGGGAAGAAATGAAGAAGT 1786
Db 6496 AAAATTAGTGTATCTCTGAACAGATAAATTTATTTGTTTCAGTGGGAAGAAATGAAGAAAGT 6555
Qy 1787 ATATCTTAAATCTATTAATCTTGGGAGTATTTTAAATGATTTTGTAAACAGTAGACT 1846
Db 6556 TTATCTTAAGGAATTAATTTAGGAGGAATTTTAAATGATTTTGTACAGTGGGACA 6615
Qy 1847 CATGAACACAGTTTAAATCGATTGATAAAGAGATTGATTTATTTGAAAAAATGGAAGT 1906
Db 6616 CATGNACAGAGTTTCAACCGTCTTATTAAGAGTTGATAGATTAAAGGACAGGTGCT 6675
Qy 1907 ATAACCGACGAATAATTTATTTCAACAGAGATATTTCTGACTATTTTCCAGATATTTGCAAG 1966
Db 6676 ATTGATCAAGAAGTGTTCATTTCAACAGGTTTACTCAGACTTCGAACCTCAGAAATTTGCAG 6735
Qy 1967 TATAAAAAATTTCTCAGTTACAAAGAAATGGAACAATATATTAACAATCAGAAGTAGTT 2026
Db 6736 TGGTCAAAAATTTCTCTCATATGATGATATGAATCTTCTTACATGAAGAAGCTGAGATTGTT 6795
Qy 2027 ATTTGCCACGAGGCGCCGCTACTTTTATGAATTCATTTATCAAAAGGAAAAAACAATTA 2086
Db 6796 ATCACACATGCGGCCCGCCGAGCTTTATGTCAGTTATTTCTTAGGGAATTTACAGTT 6855
Qy 2087 TTGTTTCTAGACAAAAAAGATGTTGGTGAACATGTAATGATCAATCAAGTAGAGTTTGTGA 2146
Db 6856 GTTGTCTCTAGAGAAAGACAGTTTGTGTGAACATATCAATGATCATCAATACAAATTTTA 6915
Qy 2147 AGAAGAAATTTTCAAGATAATAATATTTATTTATAGAAAATATAGATGATTTGTTGAA 2206
Db 6916 AAAAAATTTGCCACCTGTATCCCTTGGCTTGGATTGAAGATGATAGTGGACTTTCGCGAA 6975
Qy 2207 AAAATTATGAAGTTTCTAAGCAACTAATTTTACATCAAAATAAATTTTGTGCA 2266
Db 6976 GCGTTGAAAGGAATATAGCTACAGAAAAATATCAGGGAATAATGATATGTTTGTGAT 7035
Qy 2267 AGATTAAACAAATAGTTTCAAAAAAT 2292
Db 7036 AAAATTAGAAAAATATAGGTGAAT 7061
```

```
ID XX ADF43363 standard; DNA; 17276 BP.
AC XX ADF43363;
DT DT 12-FEB-2004 (first entry)
DE DE Streptococcus capsular polysaccharide DNA seq id 83.
KW KW receptor; lipid-based tumour associated antigen; cytostatic;
KW KW antimicrobial; gene therapy; neoplastic disease; tumour; cancer;
KW KW infectious disease; capsular polysaccharide; ds.
OS OS Streptococcus.
XX XX US2003157113-A1.
XX XX 21-AUG-2003.
XX XX 28-DEC-2000; 2000US-007511708.
XX XX 28-DEC-1999; 99US-0173371P.
XX XX (TERM/) TERMAN D S.
XX XX Terman DS;
XX XX WPI; 2003-787326/74.
XX XX P-PSDB; ADF43364.
XX XX New receptor in a mammalian cell that inhibits regular activation by
XX XX receptors specific for lipid-based tumor associated antigens, useful for
XX XX treating a neoplastic disease or tumor, and infectious diseases.
XX XX Example 3; SEQ ID NO 83; 151pp; English.
XX XX The invention describes a receptor in a mammalian cell that inhibits
XX XX regular activation by receptors specific for lipid-based tumour
XX XX associated antigen. The receptor has cytostatic and antimicrobial
XX XX properties and is suitable for use in gene therapy. The receptors,
XX XX methods and compositions are useful for treating a neoplastic disease or
XX XX tumour (cancer), and infectious diseases. This sequence represents a
XX XX streptococcal capsular polysaccharide polynucleotide, a cell surface
XX XX moiety, the DNA of which can be transfected into a cell with superantigen
XX XX DNA to generate antitumour immunity.
SQ SQ Sequence 17276 BP; 5953 A; 2414 C; 3043 G; 5866 T; 0 U; 0 Other;

Query Match 10.4%; Score 726.8; DB 10; Length 17276;
Best Local Similarity 59.2%; Pred. No. 9e-89;
Matches 1366; Conservative 0; Mismatches 902; Indels 38; Gaps 6;

Qy 16 TGGCATTTATTTGATGATAGCAGTTGCAATTTCTGCAATCTTAACAAGTCATATACCAA 75
Db 4765 TAGCGATTTCAACACAGTTGATGTTATTTTCTGCAAGTTTGACATTAACAATTAATTA 4824
Qy 76 ATGCTGATTTAAATTCG---TTCTGCAATTTTATCATAATGATGTTCAATTTTTCGAT 132
Db 4825 CTCCCAATTTTAAAGCAATAAAGATTTATTTGTTGTTCTATTGATACATATATTTT 4884
Qy 133 TTTTATATCTCGTATGCGCAGTTGAAATTTGAGTATAGAGGTAATCTGATAGAGTTTGAAA 192
Db 4885 TTTATCTTCTGATTTTACAGAGACTTTTGGAGTCGTGGCTATCTTGAAGAGTTTAAAA 4944
Qy 193 AAACATTTAATATAGTATATATTTTGCATTTTCTACGGCAGTATCATTTTGTGG 252
Db 4945 TGGTATTTGAATACAGCTTTTACTATATTTTTCATATCAAGTTCATTTTATTTTAA 5004
Qy 253 AGAATAATTTCCGCACTTTCAAGACGTCGTGCGGTGATTTTCCATATTAATAAATTCGTTT 312
Db 5005 AAACTCTTTTACACGACGACGCTTTTCTTTTACTTTTACTTTTATGCTATGATTCGATTT 5064
Qy 313 TGGTATACCTATTAAACGTAATTTAAGCAGTTTAAAGGATAGCTTTCTATTTTTCGACAA 372
```

Db	5065	TATTATATCTATTGAAATTCAATTTTAAAAATATTATCGAAATATTCTTACGCTAAGTTTTT	5126
Qy	373	TCTATCAAAAAAGACGATTCTAATTACAACGGCTGAACGATGGGAAAAATATGCAAGTTT	432
Db	5125	CACGAGATACCAAGATTGTTTGGATAAGATANGGATTTCTTTATCAAAATGACCTTTA	5184
Qy	433	TATTTGAATCACATAAACAANAATCCAAAAAATCTTTGTGCAATGGTATGTTTTAGGTACAG	492
Db	5185	GGATAAAAT---ACGACCATAATTATATCGCTGTCTGATCTTGGACTCTCTGAAAAGG	5241
Qy	493	AAATAGATATAAAATTAATTTATCATTCACGCTCTATTATTCTGTGGAGAAGCTATAGAGT	552
Db	5242	ATTGTTATGATTTGGAACATAAATCTGTTAAGGATAAATAAACAAGATGCTCTTACTTTACG	5301
Qy	553	TTTTCAACAAGGGAAGTGTGACACAGCTCTTTATAAAATCTACCAAGTGAGTTTTAGACG	612
Db	5302	AGTTAACTGCTTAACTGTGATCAGCTTTTATTAACATACCCATTGAAATTTATTTGGTA	5361
Qy	613	TAAAGCAA-----TTGCTTTTCAGATTTTGGATTTGTAGTATTGATGTAAAGCGTTG	663
Db	5362	AAATACCAAAATACAAGATATTATTAAATGACATTTGAAGCAATGGGAGTGATTGTCAATGTTA	5421
Qy	664	ATATTAAATCAATTCGGTTTTTACTCGGTTGAAAAACAACAAAAATCCAACTGCTAGGTGACC	723
Db	5422	ATGTAGAGGCATTTAGCTTTTGATTAATATAGGAGAAAAAGCGAATCCAAACTTTTGAAGGAT	5481
Qy	724	ATAGCATTTGAACTTTTTCCACAATTTTTATTAAGCCTAGTTCATATCATGATGAAGACGAC	783
Db	5482	ATAGTGTTATTAATATTCTATGAAATTTCTATAAATATAGTTCACCTTATAGCAAAACGAT	5541
Qy	784	TTTTGGATATATCTCGGACGGTAGTCGGGTTTAAATTTATTTGTGGTATAGTTTCTATTTTTGT	843
Db	5542	TTTTGGGATATCACGGTGCTATTATATAGTTTGCTCATATGTGGCATTTGTGCAATTTTTC	5601
Qy	844	TAGTTCCAAATTTATTCGTAGAGATGGTGAACGGCTATTTTTTGTCTCGAACAAGATGGAC	903
Db	5602	TAGTTCCGCAATCAGAAAAAGATGTGGACCGGCTATCTTTTCTCAAAAATAGAGTAGGTC	5661
Qy	904	AGAAATGACGCATATTTACATCTCTACAAGTTTTCCATCGATGTATGTTGATGCTCGAGAGC	963
Db	5662	GTAATGTTAGGATTTTTAGATTCTTATAATTCAGATCAATCGGAGTAGATGCAGAACAA	5721
Qy	964	GCAAAAAAGACTTCTCTCAGCCAAAAACAGATGCAAGGTTGGGTATGTTTAAAAATGGAA	1023
Db	5722	TTAAGAAAGATTTATAGTTTACAAATCAAATGACAGG-----CTAATGTTTAAGTTAGA	5776
Qy	1024	AAACGATCTAGAAATTAATCTCAATTTGGACATTTTCATACGCAAAAAACAAGTTTACAGAGT	1083
Db	5777	AGATGATCCTAGAAATTAATCAAAAATAGGAAAAATTTATTTCG-AAAAACAAGCATATAGATGAT	5835
Qy	1084	TACCACAGTTTTTATAATGTTTTTAATTCGCGATATAGTCTAGTTGGTACAGTCCACCTA	1143
Db	5836	TGCTCTCAATTTCTAATGTTTTTAAAGCGGATATAGTGTTTAGCAGGAACAGCCCTCCCA	5895
Qy	1144	CAGTTGATGAATTTGAAAAATATATCTCTGCTCAAAAAGACGATTCAGTTTAAAAACCG	1203
Db	5896	CAGTTGATGAATATGAAGAATATAATTCACGCGAAGCGACGCTTAGTTTAAAGCCAG	5955
Qy	1204	GGATTACAGGTCTCTGGCAGGTTTAGTGGTCTAGTATAATACAGACTTCGACGACGTAG	1263
Db	5956	GAATCACTGGTTTCTGGCAAAATATCTGGTAGAAAATAATTTACTGATTTTGTGATGAAATCG	6015
Qy	1264	TTCCGTTGGACTTAGCATACATTTGCACTATCTGTCAGATATTAAAAATTTTAT	1323
Db	6016	TAAAGTTAGATGTTTCAATATATCAATGAATGGTCTATTGGTCAGATATTAAAGATTATTC	6075
Qy	1324	TAAAGACAGTCAAAAGTTGTAATTTGTTGACAGAGGGAAGTAAGTAAAAAGTA-----	1372
Db	6076	TCCTAAACACTAAAGGTAGTTTACTCGGACAGGAGCTAAGTAAAGGTNAGGTTTGAAG	6135
Qy	1373	-----TATGAAAGTTTGTGTTGGTTCGGTTCTTTCAGGGGGAACATTTGATCTCATTTGATTTT	1426
Db	6136	GAATATAATGAAAAATTTGCTCTGGTTGGTTTCAAGTGGTGGTCAATCTAGCACACTTGAACCT	6195

RESULT 9  
ADO40235

ADO40235  
ID ADO40235 standard; DNA: 25020 BP.

AA  
AC  
ADQ40235;

22 XX

DT 15-JUL-2004 (first entry)

XX  
DE S. agalactiae capsular polysaccharide synthesis (cps) gene.

XX Group B Streptococci; GBS; capsular polysaccharide synthesis; cps;  
KW

XX  
DE S. agalactiae capsular polysaccharide synthesis (cps) gene.

XX Group B Streptococci; GBS; capsular polysaccharide synthesis; cps;  
KW



1373 -----TATGAGAGTTTGTGGTGGTCTTTCAGGGGACATTTTGACCTCATTGTTATT 1425  
 6525 GGAATATAAGAAATTTGCTGGTGTCAAGTGGTGGTCACTAGCACACTTTGAACC 6584  
 1426 TGTATAAACCGTTTGGGAAGGAAGAAAGCTTTTGGGTAAACATTTTGATAAAGAGGATG 1485  
 6585 TTTTGAACCAATTTGGGAAATAAAGATAGTTTGGGTAAACCTTTTGATAAAGAGATG 6644  
 1486 CAAGAAGTCTTTTGAAGAATGAAGAAATGATATCATGTTTCCAAACAAATCGCAATC 1545  
 6645 CTAGGAGTATTCTTAAGAGAAGAGATGCTATATCATGTTTCCAAACAAACCGTAATG 6704  
 1546 TCATTAAATTTAGTGAAGAAATCTTCTTAGCTTCAAAATTTTACGTGATGAGAAACGAG 1605  
 6705 TCAAAACCTTGGTAAATAAATCTATCTAGCTTTTAAGGCTTTAGAAAGAAAGACGAG 6764  
 1606 ATGTTATTATTTCATCTGGTGGCGCGTGTGCTGCCCTTCTTTTACATCGGAAACATAT 1665  
 6765 ATGTTATCATATCATCTGGTGGCGCTGTAGCAGTACCATCTTTTATATTGTAAGTTAT 6824  
 1666 TTGAGCAAGAGAGATTTATTTGAAGATTTTGAATCGAGTTAATAATCTACATTAATCG 1725  
 6825 TTGTTGTGAAGCCGTTTATATAGAGGTTTTCGACAGGATAGATAAAACCAACTTTGACAG 6884  
 1726 GAAACTAGTTTATCCCGTAACAGATATTTTATTGTTTCAGTGGGAAGAAATGAGAGG 1785  
 6885 GAAATTTAGTATCTCTGTAAACAGATAAATTTATTGTTTCAGTGGGAAGAAATGAAAGG 6944  
 1786 TATATCTTAATCTATTAACTTTGGGAGTATTTTAAATGATTTTGTAAACAGTAGGAAC 1845  
 6945 TTTATCTTAGGCAATTAATTTAGGAGAAATTTTAAATGATTTTGTCAAGTGGGAC 7004  
 1846 TCATGAACACAGTTTAAATCGATTGATAAAGAGATTTGATTTATTTGAAAAAATGGAAG 1905  
 7005 ACATGAACAGCAGTTCAACCGTCTTATTAAGAGAGTTGATAGATTAAAGGGACAGGTGC 7064  
 1906 TATAACGCAAGAAATATTATTCACAGGATATCTGACTATATTTCCAGATATTCCAA 1965  
 7065 TATTGATCAAGAGTGTTCATTCACAGCGGTTACTCAGACTTTGAACCTCAGAAATTGCA 7124  
 1966 GTATAAAAAATTTCTCAGTTTACAAAGAAATGGAACAATATATTAACAAATCAGAAAGTAG 2025  
 7125 GTGGTCAAAATTTCTCATATGATGATATGAACCTTTCATGAAAGAGCTGAGATTGT 7184  
 2026 TATTGGCAGGAGCGCCGCTACTTTTATGAATTCATTTATCCAAAGGAAAAAACAAT 2085  
 7185 TATCACACATGGCGTCCAGCGAGCTTTATGAATGCCAGTTTCTAAAGGGAAAAAACAAT 7244  
 2086 ATTGTTCCCTAGACAAAGAGTATGGTGAACATGTAATCATCATCAAGTAGAGTTGT 2145  
 7245 TGTGGTTCCTAGACAAAGACAGTTTGGAGAGCATGTGAATTAATCATCAGGTGGATTTTT 7304  
 2146 AAGAAGAAATTTTACAAGATAATAATTTTATTTATAGAAAAATATAGATGATTTGTTGA 2205  
 7305 GAAAGAGTTATCTTGAAATTTGAATAGATATATTTTGAATATCATGTAATAGAGAA 7364  
 2206 AAAAATTATGAAGTTTCTTAAGCAAAAC---TAACTTTACATCAAAATAATATTTTTTTG 2262  
 7365 TATTATTAAAGGAAAAAATAATATCTACTAGTAAAGTAAATATATCAAAAACATGATTTTG 7424  
 2263 TGAAGATTAAACAAATAGTTGAAATAATTTAATGAGGATCAAGAAATGATTAATAAAA 2322  
 7425 TTTCTCTTTCAAAATGAACATTTTCATAAACTATTTTGAATAAATATATTTTGTGGAGAA 7484  
 2323 AAGATGCATATTTTGATA 2339  
 7485 AAAAATTGAATTAACA 7501

RESULT 10  
 ADM79774  
 ID ADM79774 standard; DNA; 2226 BP.  
 XX

AC ADM79774;  
 XX 03-JUN-2004 (first entry)  
 XX Group B Streptococcus cpsD-cdsE-cpsF-cpsG partial consensus DNA sequence.  
 DE group B streptococcus; GBS bacterium; cpsD; cpsE; cpsF; cpsG; cpsL gene;  
 KW M gene; GBS infection; ds.  
 KW Streptococcus sp. 'group B'.  
 OS WO2003025216-A1.  
 PN 27-MAR-2003.  
 XX 18-SEP-2002; 2002WO-AU001281.  
 XX 19-SEP-2001; 2001AU-00007749.  
 XX (WSYD-) WESTERN SYDNEY AREA HEALTH SERVICE.  
 PA Fanrong K, Gilbert G;  
 PI WPI; 2003-381495/36.  
 XX Typing a group B streptococcus (GBS) bacterium for diagnosing the GBS  
 PT infections in pregnant women, elderly or immunocompromised patients by  
 PT analyzing the sequence of the regions in the cpsD, cpsE, cpsF, cpsG or  
 PT cpsL/M gene of the bacterium.  
 XX Claim 2; Fig 1; 106pp; English.  
 PS This invention relates to a novel method of typing a group B  
 CC streptococcus (GBS) bacterium which comprises analysing the nucleotide  
 CC sequence of one or more regions within the cpsD, cpsE, cpsF, cpsG and/or  
 CC cpsL/M genes of the bacterium, where the regions comprise one or more  
 CC nucleotides having sequences that vary between types. The method is  
 CC useful for preparing a composition for serotyping and/or subtyping a GBS  
 CC bacterium for diagnosing GBS infections in pregnant women, elderly and/or  
 CC immunocompromised patients. The present sequence is that of a consensus  
 CC DNA sequence for the 3' end of the cpsD-cpsE-cpsF and the 5' end of the  
 CC cpsG sequences of group B Streptococci which is related to the method of  
 CC the invention.  
 XX SQ Sequence 2226 BP; 767 A; 293 C; 417 G; 749 T; 0 U; 0 Other;  
 Query Match 9.3%; Score 649.2; DB 11; Length 2226;  
 Best Local Similarity 59.5%; Pred. No. 2.5e-78;  
 Matches 1223; Conservative 0; Mismatches 793; Indels 38; Gaps 6;  
 QY 16 TGGCATTTATTTGATATGATAGCAGTTGTCGAATTTCTGCAATCTTAAACAAGTCATATACAA 75  
 DB 182 TAGCGATGATTCAACACAGTTGTGGTTATTTTCTGCAAGTTTGCACATTAAACATTAATTA 241  
 QY 76 ATCTGATTTAATCG---TTCTGGAATTTTATCATATGATGTTTCATTATTTTCAT 132  
 DB 242 CTCCCAATTTTAAAGCAATAAAGATTTATTTGTTGTTCTATTGATACATATATTTGCT 301  
 QY 133 TTTTATATCTGATGCCAGTTTGAATTTGAGTATAGAGTAAATCTGATAGAGTTTCAAA 192  
 DB 302 TTTATCTTCTGATTTTACAGAGACTTTTGGAGTCGTGGCTATCTTGAAGAGTTTAAAA 361  
 QY 193 AAACATTTAACTATAGTATATAATTTTGCATTTTCTTACGGCAGTATCATATTTTGTGG 252  
 DB 362 TGGTATTGAATACAGCTTTTACTATATTTTTCATATCAAGTTCATTATTTTATTTTAA 421  
 QY 253 AGAATAATTTTCGCACTTTCAAGAGCGTGGTCCCGTATTTTACATTAATAAATCTCGTT 312  
 DB 422 AAAACTCTTTTACAAACGACACGACTTTCCTTTTACTTTTATTTGCTATGATTCGATTT 481  
 QY 313 TGGTATACCTATTAACTGTAATTTTAAAGCAGTTTAAAGGATAGCTTTCTATTTTTCGACAA 372  
 DB 482 TATTATATCTATTGAATTCATTTTTAAATAATTTATATCGAAAAATATTTCTACGCTAAGTTT 541









Db 861 GATTTCGAAAGGATGGGGATCTGCTATTTTGTCTCAGAGCGGTATAGGAAAAAATGGTCG 920  
Qy 914 CATATTTACATTCACAAATTTTCATCGATGATGTTGATGCTCAGAGCGGCAAAAAAGA 973  
Db 921 TCAGTTTCACTTTTATAGTTTCGCTCTATGTTGTAGATGCCGAGCGGCAAAAAAGAGA 980  
Qy 974 CTTCCTCAGCCCAAAACAGATGCAAGGTTGGTATGTTTAAATGCGAAAAACGATCCT 1033  
Db 981 ACTCATGGAACAAATACCATGCAAGGTGGAAATGTTT-----AAGGTGGAGATGATCCT 1035  
Qy 1034 AGAATTACTCCAATGGACATTTTCATACGCAAAAAACAAGTTTATAGACGAGTTTACCACAGTT 1093  
Db 1036 CGTATCAGAAAATGGTTGTTTATACG-GAAGACTAGCTTGGACGAGCTACCACAGTT 1094  
Qy 1094 TTATAATGTTTAAATGGCGATGATGATGTTGGTACAGTCCACCTACAGTTGATGA 1153  
Db 1095 TTATAATGTTTAAAGGGAGATATGAGTTTGGTGGCACACGCGCCACCAACAGTGGACGA 1154  
Qy 1154 ATTTGAAAAATATATCTCTGCTCAAAAGAGACGATTCAGTTTAAACCCAGGATTACAGG 1213  
Db 1155 GTATGAGCACTATACCCAGAACAAACGCGGACTAAGTTTAAACCTGCGATAACAGG 1214  
Qy 1214 TCTCTGCGAGTTAGTGTCTGATGATATATACAGACTTTCGACGAGTGTTCGGTTTGA 1273  
Db 1215 TTTATGCGAGTCAAGCGGACGAGTGAGATCAAGAATTTTCGATGAAGTTGTCAAATTAGA 1274  
Qy 1274 CTTAGCATATGATTAATGGACTATCTGTCAGATATTAATAATTTTATTAAGACAGT 1333  
Db 1275 TGTGGCTTATATGATGTTGGACAATCTGGAAGATATTAATAATTTTATTTGAAGACAGT 1334  
Qy 1334 GAAAGTTGTATGTTGTAGAGAGGGAAGTAAGTAA 1367  
Db 1335 TAAAGTAGTATTTATGAGAGATGGAGCGAAGTAA 1368

## RESULT 13

ID ADR93668 standard; DNA; 1443 BP.

XX ADR93668;

XX 16-DEC-2004 (first entry)

XX Novel S. pneumoniae DNA sequence, SEQ ID 2303.

XX Meningitis; bacteraemia; pneumonia; otitis media; ds;  
KW bacterial infection.

XX Streptococcus pneumoniae.

XX US6800744-B1.

XX 05-OCT-2004.

XX 30-JUN-1998; 98US-00107433.

XX 02-JUL-1997; 97US-0051553P.

XX 12-MAY-1998; 98US-0085131P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm LA, Bush D;

XX WPI; 2004-697205/68.

XX P-PSDB; ADR96271.

XX New isolated nucleic acid encoding a Streptococcus pneumoniae  
PT polypeptide, useful for diagnosing, preventing and/or treating  
PT pathological conditions resulting from the bacterial infection.

XX Disclosure; SEQ ID NO 2303; 151pp; English.

XX The invention relates to an isolated nucleic acid comprising a sequence

CC encoding a Streptococcus pneumoniae ADR913662polypeptide, or its  
CC fragments, with any of 9 fully defined sequences (appearing as ADR94308,  
CC ADR9489, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682,  
CC ADR96079) or any of the fully defined sequences appearing as ADR91705,  
CC ADR91886, ADR92197, ADR92234, ADR93039, ADR93079, ADR92366, ADR92650 or  
CC ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide  
CC sequences, or at least 40, 60 or 300 consecutive nucleotides, which is  
CC hybridisable under high stringency conditions to the nucleotide sequence.  
CC The nucleic acids and proteins are chosen from 5206 disclosed sequences.  
CC Also included are a recombinant expression vector comprising the isolated  
CC nucleic acid cited above operably linked to a transcription regulatory  
CC element, a cell comprising the recombinant expression vector and a probe  
CC comprising at least 20 consecutive nucleotides of the nucleotide  
CC sequences as cited above. The methods and compositions of the present  
CC invention are useful for the diagnosis, prevention and/or treatment of  
CC pathological conditions resulting from bacterial infection by  
CC Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and  
CC otitis media. The present sequence is one of the 2603 disclosed S.  
CC pneumoniae nucleic acid sequences. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=6800744B1.

XX SQ Sequence 1443 BP; 432 A; 218 C; 340 G; 453 T; 0 U; 0 Other;

Query Match 5.9%; Score 415.8; DB 13; Length 1443;

Best Local Similarity 58.3%; Pred. No. 5e-47;

Matches 789; Conservative 0; Mismatches 552; Indels 12; Gaps 3;

Qy 14 ATTGGCAATATTGATGATGATAGCAGTTGCAATTTCTGCAATCTTAACAAGTCATATACC 73

Db 102 ATTGGCCATAGTCCAGAGTTTCTGTTATTTTATGTTGCTTATCTACTAGCACTGTGAG 161

Qy 74 AAATGCTGATTTAAATCGTTCTGGAATTTTATCATTAATGATGGTTCATATTTTTCGATT 133

Db 162 AGAAACAGAGATTGTTTCAACACAGCTATTGTACTTATATATCTCCACTATTTTGCCTT 221

Qy 134 TTTTATATCTCGTATGCCAGTTGAAATTTGATATAGAGTAACTGATAGAGTTTGAATA 193

Db 222 TTATATCAGTGATTTATGGACAGGATTTCTTTAAAGGGGATATTTGATGAACCTTATCCA 281

Qy 194 AACATTTAACTATAGTATATAATTTTGCATTTTCTTACGGCAGTATCATTTTGTGTGA 253

Db 282 GACATTTGAATATATCTTCTTGGCTAGCGATAGTATTTCTTAATTTTCTTTAGA 341

Qy 254 GAATAATTTGCGCACTTTCAAGACGTTGGTCCGTTATTTCCATTAATAAATCTTGTGTTT 313

Db 342 AGATCGATTTAGTATTTTCCAGAGAGCGCATGATTTACTTCTCTCTATTACATGTTCTCTT 401

Qy 314 GGTATACCTATTTAAGCTAATTTAAGCAATTTTAAAGGATAGTCTTCTATTTCGACAAT 373

Db 402 AGTCTATGTCTAAACCGATTTTATCAAGTGGTATTTGGAACGGGCTTATCCCAACTTTAA 461

Qy 374 CTATCAAAAAAGACGATTTCTAATTTACACCGCTGAACGATGCGAAAAATATGCAAGTTT 433

Db 462 AGGAAGTAGAAGATTTCTCTACTTACAGCAACTTCTCGTGTGCAAAAGGATTTGGATAG 521

Qy 434 ATTTGATACATATAAACAATTTCAAAAAAATCTTGTGTGCAATTTGATGTTTGTAGTACAGA 493

Db 522 ACTAATAGAATCAGATGATGTTCTGTGGGAGTTGGTAGCGCTCAGTGTGTTTGTAGATAACC 581

Qy 494 AATAGATAAAATTAATTTATCATTCACGCTCTATTATTTCTGTGGAAGAAGCTATAGAGTT 553

Db 582 AGATTTTTCAGCATGATTTATTTAAAGGTTGTAG-----CAGAGGGGGAGATCGTAAACTT 635

Qy 554 TTCAACAAAGGGAAGTGTGCGACCACTCTTTATAAATCTTACCAAGTGAGTTTGTAGACGT 613

Db 636 TGGCATCTCATGAGTGTGCGTGAAGTCTTTATCAATCTTCCAGTGAAAAATACATAT 695

Qy 614 AAAGCAATTCGTTTCAGATTTTTCAGTTTGTAGTATTTGATGTAAGCGGTTGATATTAATTC 673

Db 696 TGGAGAGCTTGTCTCTCAGTTTGAACGATGGGAGTTGATGTAACAGTCAATCTAAATGC 755







FT		/note= "encodes AAY54069"		QY	2	TGCCAAACGAAATTTGGCAATTTATTTGATATGATAGACAGTTGGCAATTTCTCGCAATCTTAAAC	61
FT	CDS	8633..10681		Db	5649		
FT		/*tag= f					
FT		/product= "EPS6"					
FT		/note= "encodes AAY54070; contains 1 stop codon at		QY	62	RAGTCATATACCAATGCTGAATTTAAATCGTTCTGGAAATTTTATCATATAATGATGGTTCA	121
FT		nucleotides 9410-9412"		Db	5709		
FT	CDS	10703..11671					
FT		/*tag= g					
FT		/product= "EPS7"		QY	122	TTATTTTGCATTTTATATATCTCGTATGCGCAGTTTGAATTTTGAGTATAGAGGTAAATCTGAT	181
FT		/note= "encodes AAY54071"		Db	5769		
FT	CDS	12838..13788					
FT		/*tag= h					
FT		/product= "EPS8"		QY	182	AGAGTTTGAAAAAACAATTAATACTATAGTATAATATTTGCAATTTTCTTACGGCAGTATC	241
FT		/note= "encodes AAY54072"		Db	5829		
FT		14138..15553					
FT		/*tag= i					
FT		/product= "EPS9"		QY	242	ATTTTGTGGAGAAATAATTTTCGCACATTTCAAGACGTGGTGGCCGTGATTTTCACATTAAT	301
FT		/note= "encodes AAY54073"		Db	5889		
FT	CDS	16919..18016					
FT		/*tag= j					
FT		/product= "EPS10"		QY	302	AAACTTTCGTTTGGTATACCTTATTTAAAGTAAATTTAAGCAGTTTAAAGGATAGCTTTCT	361
FT		/note= "encodes AAY54074"		Db	5949		
FT							
XX	W09962316-A2.						
XX				QY	362	ATTTTCGAATCTATCAAAAAAGACGATTTCTAAATTAACAACGCTGAACGATGGGAAAA	421
XX	09-DEC-1999.			Db	6009	TACACGTCGTAAGAGTAACAAGATAATTTCTCTGATTTCTGATCAAGACAGTCTAGAAAA	6068
XX	22-APR-1999;	99WO-EP002841.		QY	422	TATGCAAGTTTATTTGNAATCACAATAACAATTTCAAAAAATTTCTTGTGCAATTTGGTAGT	481
XX	22-APR-1998;	98EP-00201310.		Db	6069	T-----GTTTTGTCTCGTATGAAAAGACAATATGATGGTGGTAGGATTAACAGCAGT	6116
XX	22-APR-1998;	98EP-00201311.		QY	482	TTTAGGTACAGAAATAGATAAAATTAATTTTATCAATTACCTCCGCTCTATTATTCTGTGGAAGA	541
XX	22-APR-1998;	98EP-00201312.		Db	6117	TTGTGCTTGGATAATCCTTATTTTACCGATCCATTTTATCAAGAGTGTTTAAACCTGAAAA	6176
XX	(NEST ) SOC PROD NESTLE SA.			QY	542	AGCTATAGAGTTTTCACAAAGGGGAAGTGGTCGACACAGTCTTTTATAAAATCTACCAAGTGA	601
XX	Stingele F, Germond JE, Lamothe G;			Db	6177	TTTGATTGAATATGCGACACACTCAGTAGTAGACCAAGTTTGTGATTAATCTGCCAAGTGG	6236
XX	WPI; 2000-097267/08.			QY	602	GTTTTTACACGTAAGCNAATTCGTTTTCAGATTTTGAGTTTGTAGGTATTGATGTAAGCGT	661
XX	P-PSDB; AAY54065, AAY54066, AAY54067, AAY54068, AAY54069, AAY54070,			Db	6237	GCAGTATAGAAATTTGGGATTATGCATCACCTTTTGGAGATCATGCGGAATTTCCAGTTTCTAT	6296
XX	AAY54071, AAY54072, AAY54073, AAY54074.			QY	662	TGATATTAATTCATTTCCGTTTCTGCGTTGAAAAACAAAAAATCAACTGCTAGGTGA	721
XX	New recombinant enzymes for synthesis of exopolysaccharides, particularly			Db	6297	TAATTTGAAATGCCCTTGAATTTTATGAATCAAGGTGAAAAACGTAATTCACAAATGGGTCC	6356
XX	in lactic acid bacteria, for improving properties of fermented milk			QY	722	CCATAGCATTTGTAATCTTTTCCACAAATTTTATAAGCCTAGTCATATCATATGATGAAACG	781
XX	products.			Db	6357	TTTCAAGTTGTTAGCTTTTCAACGTAATTTTATAGCTATGGAGATATCTTGGCGAAACG	6416
XX	Claim 6; Page 77-96; 162pp; French.			QY	782	ACTTTTGGATATATCTCGAGCGGTAGTCGGGTTAATTTATTTGGTGTATAGTTTCTATTTT	841
XX	The present sequence encodes enzymes involved in the biosynthesis of			Db	6417	TTTCTCGATATCTGTGGAGCTCTAGTTGGTTTGTCTCTGTGGTATTTGTAGGAATCTT	6476
XX	exopolysaccharides (EPS). These enzymes are designated EPS1-EPS10. and			QY	842	GTTAGTTCCAATTAATTCGTACAGATGTTGGACCGGCTATTTTCTCTCAGAAAAACGATTTG	901
XX	are encoded by open reading frames eps1-eps10. The enzymes are isolated			Db	6477	CTTTTATCTCTTATTCGTAAGATGGAGGACCAAGCCATTTTCTCTCAAGCCGTGTGGG	6536
XX	from Streptococcus thermophilus strain Sfi39. The proteins are used in a			QY	902	ACAGAAATGGACGATTTTATTAATTTCTACAAGTTTTCGATCGATGTATGTTGATGCTGAGGA	961
XX	method for the synthesis of EPS, which includes at least one step of			Db	6537	AGAAATGGACGATCTTCTCAAGTTTATAAATCCGTTCTATGTGTGTGTGATGCGGAAGA	6596
XX	forming a bond (alpha or beta-isomer) between C-1 (carrying the reducing			QY	962	GGGAAAAAGACTTGTCTACGCCAAAAACAGATGCAAGGGTGGGTATGTTTTTAAATGGG	1021
XX	aldehyde function, of an activated D-galactose pyranose), and a phosphate			Db	6597	AATCAAGAAGAATTTGATGGCACAGAAATCAAATGCTGTGGTGTATGTTT-----AAGATG	6651
XX	on a lipophilic or proteinaceous primer. Stepwise synthesis of EPS occurs			QY	1022	AAAAACGATCTCTAGAAATTTACTTCCAAATTTGGACATTTTCATAGCAAAAAACAAGTTT	1081
XX	with, in each step, addition of a new sugar unit, through its hemi-acetyl			Db	6652	GCAATGATCCACGTAATTTACCAAAATTTGGACATTTTCAATTCG-THAAAACGAGTCTT	6710
XX	function, to an alcoholic hydroxyl of a second sugar unit, present at the			QY	1082	GTTACCAACGATTTTATTAATGTTTAAATTTGGCGATATGATGTTAGTTTGGTACACGTT	1141
XX	end of a chain of sugar residues bonded to the primer. EPS1 to EPS4 are			QY			



Db 6711 ACTTCCACAATTTGGAAATGTTCTAAAGGTGATATGAGCTTGGTGGGACACGTCTCTCC 6770  
Qy 1142 TACAGTTGATGCAATTTGAAAAATATCTCTCTGGTCAAAAGAGACGATTTGAGTTTAAACC 1201  
Db 6771 AACAGTTGATGATGCAAAATATACACCTGAACGAAACGTCGTTAAAGTTTAAACC 6830  
Qy 1202 AGGATTTACAGGTCCTCGGACAGGTAGTGTGCTAGTAAATATACAGACTTCGACGACGT 1261  
Db 6831 TGGTATCACCTGGCTTTTGGCAAGTAAGCGGTGCAAGTGAATTTACTGATTTGATGAAGT 6890  
Qy 1262 AGTTCGGTTGAGCTTACATACATATGATATTTGAGAGAGGGAAGTAAGTAAAGATATATGAAGT 1321  
Db 6891 TGTAAGACTAGACGTTGCTTATTGGACGGATGACCAATCTGGCGTGATATCAAAATCTT 6950  
Qy 1322 ATTAAGACAGTGAAGTCTGATTTGTTGAGAGAGGGAAGTAAGTAAAGATATATGAAGT 1381  
Db 6951 ATTGAAGACGATTAAAGTAGTAGTAAAGAGGATGAGGAAAGTGGATGCGCTTCCACCAT 7010  
Qy 1382 TTGTTTGGTCG 1392  
Db 7011 TCTTTTAATGG 7021

## RESULT 17

ABAO1447  
ID ABAO1447 standard; DNA; 28170 BP.

AC ABAO1447;

DT 21-FEB-2002 (first entry)

XX Streptococcus thermophilus eps10 operon #2.

DE Exopolysaccharide; lactic acid bacterium; eps; fermented food product;  
KW ds.

XX Streptococcus thermophilus.

OS W0200179500-A2.

XX 25-OCT-2001.

XX 18-APR-2001; 2001WO-FR001199.

XX 18-APR-2000; 2000FR-00004972.

XX (INRG ) INRA INST NAT RECH AGRONOMIQUE.  
PA (DANO-) CIE DANONE SA GERVAIS.  
PA (RHOD ) RHODIA CHIM.

XX Rallu F, Bésancon-Yoshpe I, Fremaux C, Mengaud J, Renault P;

XX WPI; 2002-017616/02.

XX New nucleic acid fragments containing exopolysaccharide operon, useful  
PT e.g. for increasing exopolysaccharide synthesis in lactic acid bacteria.

XX Claim 9; Page 109-117; 144pp; French.

XX The present sequence is an eps operon from Streptococcus thermophilus.  
CC Proteins encoded by the eps operon function in exopolysaccharide (EPS)  
CC synthesis. The operon is useful for producing chimeric eps operons, for  
CC optimising production of EPS in lactic acid bacteria. EPS impart texture,  
CC mouth feel and rheological properties to fermented food products (e.g.  
CC yoghurt). They function as thickeners, to provide free-flowing and creamy  
CC texture, and may also have biological activities beneficial to health

XX Sequence 28170 BP; 8951 A; 4464 C; 5468 G; 9287 T; 0 U; 0 Other;

XX Query Match 5.4%; Score 376; DB 6; Length 28170;

XX Best Local Similarity 56.7%; Pred. No. 9.7e-42;

XX Matches 782; Conservative 0; Mismatches 580; Indels 18; Gaps 4;

Qy 13 AATTGGCAATTTATGATATAGCAGTTGCAATTTCTGCAATCTTAAACAAGTCATATAC 72  
Db 4600 AAATTGGTATTAACAGTTGGTTGGTTGTTTTCAGCCATGGTAGCTAGTAAAAATAC 4659  
Qy 73 CAATCTGATTTAAATCGTTCTGGAATTTTATCATATATGATGCTTCATATTTTGCAT 132  
Db 4660 CTTATACAGAGATTAACCAAGGAAGCATTTGCTCTTTAGGTGCTGATACATGTAGTGTCTT 4719  
Qy 133 TTTTATATCTCTGATGCCAGTTGAAATTTGAGTATAGAGGTAATCTGATAGAGTTTGAAA 192  
Db 4720 TCTATATCAGTAGTTATTAAGAAAATCTTAAGTATAGAGGCTACTTGGATGAATCATTTG 4779  
Qy 193 AACATTTAATATATATTAATTTTGCATTTTCTTACGGCAGTATCAATTTTGTGTGG 252  
Db 4780 CAACTGTCAAAATATGTTTCATATTTGCTCTAAATGTAATCTTCTCCACGCTTTTTCGAG 4839  
Qy 253 AGAATTAATTTTCGACACTTTCAAGACGTTGGTCCGCTGATTTTCAATTAATAAATCTTCGTTT 312  
Db 4840 ATGGAAGTTTCTCTATCTCAGTCGCGGACTTTTATGTTATCTGATTTTCAGGTGTTTC 4899  
Qy 313 TGGTATACCTTATTAAAGTAAATTTAAGCAGTTTAAGGATAGCTTCTATTTTCGACAA 372  
Db 4900 TCTTATACGTTACAAATATCTGTTCTTAAGTATTT-----CCGCTCTCTATCTATACAC 4953  
Qy 373 TCTATCAAAAAAGACGATTTCTAAATTAACAACGCTGAACGATGGGAAAATATCAAGTTT 432  
Db 4954 GTCGTAAAAAGCAATAAGATATTTCTCTGATTTCTGATCAAGCAGCTGTAGAAAATGTTT 5013  
Qy 433 TATTGATACATATAAACAATTTCAAAAAATCTTGTGATTTGGTAGTTTGTAGTTACAG 492  
Db 5014 TGTCTCGTATGAAGACAATATGATGGTAGGATTAAGCCGTTTGTGTCTGGATATC 5073  
Qy 493 AATATAGATAAATTAATTTTATCAATACCGCTCTATTATTCTGTGGGAAGAGCTATACAGT 552  
Db 5074 C-----TTATTTTACCGATCCATTTATCAAGAGTGTAAACCTGAAAATTTGATGAAT 5127  
Qy 553 TTTCAACAAGGGAAGTGGTCGACACGCTCTTTATAAATCTTACCAAGTGGAGTTTGTAGACG 612  
Db 5128 ATCGACACACATCAGTAGTAGACCAAGTTTGTATTAATCTTCCAAAGTGACAGTACAAGA 5187  
Qy 613 TAAAGCAATTCGTTTCAGATTTTGGATTTGTAGTATTTGATGTAAGGTTGATATTAAT 672  
Db 5188 TTTGGGATTTACGATCACTTTTGAAGTTTATGGGGATTTCCAGTATCCATTAATTTGAATG 5247  
Qy 673 CATTCGTTTACTGCTGTTGAAAACAACAAATCCAACTGCTAGTGACCATAGCATTTG 732  
Db 5248 CCTCCAAATTTATGAGTAAGGTTGAAAACGATATCCAAATTTGGGTCTCTTCAAAGTTG 5307  
Qy 733 TAACTTTTTCACAAATTTTATAAGCTTAGTCAATATCATGATGAACAGCTTTTGGATA 792  
Db 5308 TTACGTTTTCACGCAATTTTATAGCTATGGAGATATCTTGGCGAAACGTTTCTCGATA 5367  
Qy 793 TACTCGAGCGGTAGTCGGGTTAAATTTATTTGCGTATAGTTTCTATTTTGTAGTTCCAA 852  
Db 5368 TCTGTGGAGCCCTAGTTGGTTGGTCTCTGTGGGATTTGTGGAATCTTCTTTATCCAC 5427  
Qy 853 TTATTCGTAGAGATGTTGGACCGGCTATTTTCTCAGAACAGATTTGGACAGATGGAC 912  
Db 5428 TTATTCGTAGGATGTTGGGCCAGCCATTTTCTCAAGACCGTGGGGAAGAAATGGAC 5487  
Qy 913 GCATATTTTACATTTCAAGTTTTCGATGATGATGTTGATGCTGAGGAGCGCAAAAAG 972  
Db 5488 GTATCTTCAAGTTTATTAATTTCCGTTCTATGTTGTTGATGCGGAGAAATCAAGAGA 5547  
Qy 973 ACTTGCTCAGCCAAAACAGATGCAAGGGTGGGTATGTTTAAAAATGGGAAAAACGATCC 1032  
Db 5548 ATTTGATGGCAGAGATCAAAATGCTGTTGTTATGTTT-----AAGATGCAATGATCC 5602  
Qy 1033 TAGAATTTACTCCAATTTGGACATTTTATACGCAAAAACAGTTTGTAGACGAGTTACCACAGT 1092  
Db 5603 ACGTATTTACCAAAATTTGGACATTTTCATTCG-TAAAAACGAGTCTTGTGAACTTCCACAAT 5661





Db 5657 TTTGGAATGTTCTAAAAGGTGATATGAGTCTGGTAGGAACACAGCTCCACCAACATGGATG 5716  
Qy 1153 AATTTGAAAATATATCTCTGGTCAAAAAGAGACGATTTGAGTTTAAACCGAGGATTTACAG 1212  
Db 5717 AGTACGAATCTTATACACCGGAACAAAACGTCGCCTCAGCTTTAAACCGAGGTATTACTG 5776  
Qy 1213 GTCTCTGGCAGGTTAGTGGTCTAGTAATATACAGAGCTTCGACGAGGTAGTTCCGGTTGG 1272  
Db 5777 GTCTTTGGCAAGTAAGCGGTGGAAGTGAATATCTGATTTTGGATGAAGTTGTAAACTAG 5836  
Qy 1273 ACTTAGCATACATTTGATTAATTTGGACTATCTGGTCAGATATTTAAATTTTAAAGACAG 1332  
Db 5837 ACGTTGCTTATTGGACGGATGGACATCTGGCGGATATTAATCTTATTGAACAA 5896  
Qy 1333 TGAAGTTGATTTTGGAGAGAGGGAAGTAAGTAAAGTATATGAAGTTTGGTGGTCG 1392  
Db 5897 TTAAGTAGTAGTAATGAAGGATGGACAAAGTATGGCTTCTCCATTTCTTTAATGG 5956

RESULT 20  
ABA01444  
ID ABA01444 standard; DNA; 18488 BP.  
XX ABA01444;  
XX  
XX  
DT 21-FEB-2002 (first entry)  
XX  
XX Streptococcus thermophilus eps6 operon #2.  
XX  
XX Exopolysaccharide; lactic acid bacterium; eps; fermented food product;  
XX ds.  
XX  
XX Streptococcus thermophilus.  
XX  
XX WO200179500-A2.  
XX  
XX 25-OCT-2001.  
XX  
XX 18-APR-2001; 2001WO-FR001199.  
XX  
XX 18-APR-2000; 2000FR-00004972.  
XX  
XX (INRG ) INRA INST NAT RECH AGRONOMIQUE.  
XX (DANO-) CIE DANONE SA GERVAIS.  
XX (RHOD ) RHODIA CHIM.  
XX  
XX Rallu F, Besancon-Yoshpe I, Fremaux C, Mengaud J, Renault P;  
XX  
XX WPI; 2002-017616/02.  
XX  
XX New nucleic acid fragments containing exopolysaccharide operon, useful  
XX e.g. for increasing exopolysaccharide synthesis in lactic acid bacteria.  
XX  
XX Claim 9; Page 94-99; 144pp; French.  
XX  
XX The present sequence is an eps operon from Streptococcus thermophilus.  
XX proteins encoded by the eps operon function in exopolysaccharide (EPS)  
XX synthesis. The operon is useful for producing chimeric eps operons, for  
XX optimising production of EPS in lactic acid bacteria. EPS impart texture,  
XX mouth feel and rheological properties to fermented food products (e.g.  
XX yoghurt). They function as thickeners, to provide free-flowing and creamy  
XX texture, and may also have biological activities beneficial to health  
XX  
SQ Sequence 18488 BP; 5800 A; 2982 C; 3655 G; 6051 T; 0 U; 0 Other;

Query Match 5 3%; Score 371.4; DB 6; Length 18488;  
Best Local Similarity 55.9%; Pred. No. 4.1e-41;  
Matches 777; Conservative 0; Mismatches 596; Indels 18; Gaps 3;  
Qy 2 TCGCAACACGAAATGGCATTATTGATATGATAGCAGTTTGCATTTCTGCAATCTTAAAC 61  
Db 4589 TCGTCGCATTGAAATTTGATATTATACAGTTGGTTGGTTGTTTTCGCAGCATTGGTAGC 4648

Qy 62 AAGTCATATACCAAAATGCTGATTTTAAATCGTCTCGAAATTTTATCATATGATGGTTCA 121  
Db 4649 TAGTAAATATACCATATACAGAGANTTACCAGAGAGTATTTGCTCTTTTAGGTGCTGACA 4708  
Qy 122 TTATTTTGGCATTTTATATCTCGTATGCCAGTTGAATTTGAGTATAGAGGTAATCTGAT 181  
Db 4709 TGTAGTGTCTTACTATATCAGTAGTTATTATGAAAAATATTAACTATAGAGGCTACTTGA 4768  
Qy 182 AGAGTTTGAAGAAAACATTTAACTATATAGTATATATTTGCAATTTTCTTACGGCAGTATC 241  
Db 4769 TGAATTCATTTGCAACTGTCAAAATATTGTTTCATATTTGCTTAATTTGCAACTTTTCTCTC 4828  
Qy 242 ATTTTGTGGAGAAATTTTCGCACTTTCAAGCGTGGTGGCGTGTATTTTCAATTAAT 301  
Db 4829 GTTTTGTGAGATGGGAAGTTTCTATCTCACGTCCGGACATTTCTTACGTCAACCTGAT 4888  
Qy 302 AAACCTTGTGTTGGTATACCTATTTAAACGTAATTTTAAAGCAGTTTAAAGGATAGCTTTCT 361  
Db 4889 TTCAGGTGTTCTCTTATAGCTTACAATACTGTTCTTAAGTATTTCCGTTCACTATTTA 4948  
Qy 362 ATTTTCGCAATCTATCAAAAAAGACGATTTCTAAATTACAACGGCTGAACGATGGGAAA 421  
Db 4949 TACACGTCGTAAGTAAGTAACAAGAAATATTCTCTTTGATTTCTGATCAAGCACGCTCTAGAAA 5008  
Qy 422 TATGCAAGTTTATTTGAATCACATAAACAATTTCAAAAAAATCTTGTTCATTTGGTAGT 481  
Db 5009 TGTTTTGTCTTGTATGAAAGACAATATGATGGTAGGATTA-----CAGCAGT 5056  
Qy 482 TTTAGGTACAGAAATAGATAAAATTAATTTATCATTTACCGCTCTATTATTCTGTGGAAGA 541  
Db 5057 TTGTGTTGGATAATCTCTTATTTTACCGATCAATTATCAAGAGTGTAAACCTGAAA 5116  
Qy 542 AGCTATAGAGTTTCAACAAAGGAAGTGTGCGACACGCTTTTATATAATCTACAAGTGA 601  
Db 5117 TTTGATTTGAATATGCGACACACTCAGTAGTAGACCAAGTTTGTGATTAATTTGCGGAGTGA 5176  
Qy 602 GTTTTGTAGAGTAAAGCAATTCGTTTCAGATTTTGTAGTTGTAGTATTTGATCTAGCGT 661  
Db 5177 GCAGTACAAGATTTGGGATTTATGCGTCACCATTTGAACTTATGGGAATCCAGTATCCAT 5236  
Qy 662 TGATATTAATTCATTCGTTTACTCGTTGTAAGAAAACAAAAAAATCCAACTGCTAGGTGA 721  
Db 5237 TAATTTGAATGCCCTTGAATTTATGATCAGGCTGAAAACGATTCACAAATTTGGTCC 5296  
Qy 722 CCATAGCATTTGTAATCTTTTCCAAATTTTATAAGCCCTAGTCAATATATGATGAACG 781  
Db 5297 TTTCAAAAGTTGTACGTTTCTACGCAATTTTATAGCTATGGAGATATCTTGGCGAAACG 5356  
Qy 782 ACTTTTGGATATACTCGAGCGGTAGTCCGGTTAATTTATTTGTGTATAGTTTCTATTTT 841  
Db 5357 CTTCCTCGATATCTGGGAGCTCTAGTTGGTTTGGTCTCTGGTATTTGTTGGAATCTT 5416  
Qy 842 GTTAGTTTCCAAATTTTCTGATAGATGCTGGACCGCTTATTTTGTCTCAGAAAACGATGG 901  
Db 5417 CCTTTATCTCTTATTCGTAAGATGGAGACCAGCTATTTTGTCTCAAGACCGTGTGG 5476  
Qy 902 ACAGATGGACGATATTTTACATTTTCAAGTTTTCGATCGATGATGTGTGATCTGAGGA 961  
Db 5477 AGAAATGGACGATCTCTCAAGTTTATATAATTTCCGTTCTATGATGTTGATGGGAAGA 5536  
Qy 962 GCCAAAAAGACTTCTCAGCCAAACAGATGCAAGGTTGGGTATGTTTAAATGGG 1021  
Db 5537 AATCAAGAGAATTTTGTATGSCACAGAATCAAAATGTCTGGTGGTATGTTT-----AAGATG 5591  
Qy 1022 AAAAAAGATCCTAGAAATTTACTCCAAATTTGGACATTTTCATACGCAAAAAACAAGTTTAGACGA 1081  
Db 5592 GACATATGATCCAGTATATACCAAAATTTGACATTTTCAATTCG-TAAACAGAGTCTTGATGA 5650  
Qy 1082 GTTTACCAAGTTTATTAATGTTTAAATGGCGATATAGTCTAGTTGGTATACAGTCCACC 1141  
Db 5651 ACTTCCACAATTTTGGAAATGTTCTTAAAGAGGTGATATGAGCTTGTGTAGGAACAGCTCCACC 5710  
Qy 1142 TACAGTTGATGAATTTGAAAAATATATCTCTCGTGTCAAAAGAGACGATTTGAGTTTAAACC 1201





Qy	1142	TACAGTTGATGAATTTGAAAAATATACTCTGGTCAAAAGAGACGATTGAGTTTAAACC	1201
Db	5700	AACATTGGATGAGTACGAATCTTATACACGGGAACAAAAACGTCGCCTCAGCTTTAAACC	5759
Qy	1202	AGGGATTACAGGTCTCTGGCAGGTTAGTGGTCGTAGTAATATACACAGACTTTCACACAGCT	1261
Db	5760	AGGTATTACTGGTCTTTGGCAGGTAAAGCGGTGGAAGTGAATTTACTGATTTTGTGAGT	5819
Qy	1262	AGTTCCGTTTGGACTTTAGCATACATTTGATAATTGGACTATCTGTCAGATATTAAAAATTTT	1321
Db	5820	TGTAACAACCTAGACGTTGCTCTTATTGGACGGATGGACAATCTGGCGCGATATCAAAATCTT	5879
Qy	1322	ATTAAAGACAGTGAAAGTCTGTTGTTGAGAGAGGGAAGTAAGTAAAGATATATGAAAGT	1381
Db	5880	ATTCAAAACAAATTAAAGTAGTAGTAAATGAAGGATGGAGCAAGTAGTGGCTTTCACCAAT	5939
Qy	1382	TTGT	1385
Db	5940	TTTT	5943

RESULT 23

RESOL1 23  
ABA01448  
ID ABA01448 standard; DNA; 22157 BP.

DE Streptococcus thermophilus eps11 operon #2.

Exopolysaccharide; lactic acid bacterium; eps; fermented food product;  
ds.

XX PN WO200179500-A2.

XX  
PD 25-OCT-2001.

XX 18-APR-2001: 2001WO-FR001199.

XX  
PR 18-APR-2000: 2000FR-00004972.

XX (INRG ) INRA INST NAT RECH AGRONOMIQUE.  
PA (DANO-) CIE DANONE SA GERVAIS.  
PA (RHOD ) RHODIA CHIM.

XX	PI	Rallu F,	Besancon-Yoshpe I,	Fremaux C,	Mengaud J,	Renault P;
----	----	----------	--------------------	------------	------------	------------

AA  
DR  
WPI; 2002-017616/02.

AA New nucleic acid fragments containing exopolysaccharide operon, useful  
PT e.g. for increasing exopolysaccharide synthesis in lactic acid bacteria.  
PT

XX  
PS  
Claim 1; Page 117-124; 144pp; French.

The present sequence is an eps operon from *Streptococcus thermophilus*. Proteins encoded by the eps operon function in exopolysaccharide (EPS) synthesis. The operon is useful for producing chimeric EPS operons, for optimising production of EPS in lactic acid bacteria. EPS impart texture, mouth feel and rheological properties to fermented food products (e.g. yoghurt). They function as thickeners, to provide free-flowing and creamy texture, and may also have biological activities beneficial to health.

SQ Sequence 22157 BP; 7092 A; 3404 C; 4666 G; 6995 T; 0 U; 0 Other;

Query Match 5.2%; Score 364.8; DB 6; Length 22157;

Qual. match	5.28	Score 30.8, SS 6,	Length 22137,
Best Local Similarity	56.2%	Pred. No. 3.1e-40,	
Matches 775; Conservative			Indels 18; Gaps 4;
Mismatches 587;			
Mismatches 0;			

QY 13 AATTGGCATTATTGATATGATAGCAGTTGCAATTTCTGCAATCTTAACAAGTCATATAC 72

[illegible]



Db 5659 TTTGGAATGTTCTAAAGGTGATATGAGCTTGGTAGGAAACAGCTCCACCAACATTGGATG 5718  
Qy 1153 AATTGTGAATAATATACTCCTGGTCAAAAGAGACCAATTGAGTTTTAAACCAGGATTACAG 1212  
Db 5719 AGTACGAATCTTATACACCGGAACAAACACGTCGCCCTCAGCTTTAAACACGAGTATTACTG 5778  
Qy 1213 GTCTCTGCGAGGTAGTGGTGGTATGATATATACAGACTTCGACGAGCTAGTTCCGGTTGG 1272  
Db 5779 GTCTTTGGCAAGTAAGCGGTCAAGTGAATTAATCTGATTTTGTGATGAAGTTGTAAACATAG 5838  
Qy 1273 ACTTAGCATACATTGATAAATTGGACTATCTGGTCAAGATATTAAATAATTTTAAAGACAG 1332  
Db 5839 ACCTTGCTTATTTGGACGGATGGACAATCTGGCGGATATCAAAATCTTATTGAAACAA 5898  
Qy 1333 TGAAGTTGATTTCTGAGAGAGGAAGTAAGTAAAGATATATCAAAAGTTTGGTGGTCG 1392  
Db 5899 TTAAGTAGTAGTAATGAAGGATGGAGCAAGTAGTGGCTTTCCACCATTTCTTTTAATGG 5958

RESULT 24  
ABA01435  
ID ABA01435 standard; DNA; 14142 BP.  
AC ABA01435;  
XX  
XX 21-FEB-2002 (first entry)  
XX  
XX Streptococcus thermophilus eps5 operon #1.  
XX  
XX Exopolysaccharide; lactic acid bacterium; eps; fermented food product;  
XX ds.  
XX  
XX Streptococcus thermophilus.  
XX  
XX WO200179500-A2.  
XX  
XX 25-OCT-2001.  
XX  
XX 18-APR-2001; 2001WO-FR001199.  
XX  
XX 18-APR-2000; 2000FR-00004972.  
XX  
XX (INRG ) INRA INST NAT RECH AGRONOMIQUE.  
XX (DANO-) CIE DANONE SA GERVAIS.  
XX (RHOD ) RHODIA CHIM.  
XX  
XX Rallu F, Besancon-Yoshpe I, Fremaux C, Mengaud J, Renault P;  
XX WPI; 2002-017616/02.  
XX  
XX New nucleic acid fragments containing exopolysaccharide operon, useful  
XX e.g. for increasing exopolysaccharide synthesis in lactic acid bacteria.  
XX  
XX Claim 11; Page 48-52; 144pp; French.  
XX  
XX The present sequence is an eps operon from Streptococcus thermophilus.  
XX Proteins encoded by the eps operon function in exopolysaccharide (EPS)  
XX synthesis. The operon is useful for producing chimeric eps operons, for  
XX optimising production of EPS in lactic acid bacteria. EPS impart texture,  
XX mouth feel and rheological properties to fermented food products (e.g.  
XX yoghurt). They function as thickeners, to provide free-flowing and creamy  
XX texture, and may also have biological activities beneficial to health  
XX  
XX Sequence 14142 BP; 4483 A; 2605 C; 2582 G; 4411 T; 0 U; 61 Other;  
XX

Query Match 5.0%; Score 347.2; DB 6; Length 14142;  
Best Local Similarity 55.5%; Pred. No. 7.2e-38;  
Matches 766; Conservative 0; Mismatches 588; Indels 26; Gaps 4;  
Qy 13 AATTGGCAATATTGATATGATAGACAGTTGCAATTTCTGCAATCTTAAACAGTCAATATAC 72  
Db 4576 AAATTGGTATTATACAGTTGGTGGTTGTTTTTGGACGCCATGGTAGTAAATAATAC 4635

Qy 73 CAAATGCTGATTTAAATCGTTCTGGAATTTTTTATCATATGATGGTTTCATTTATTTTGCAT 132  
Db 4636 CTTATACAGAGATTACCAAGGAAGCAATTGCTCTTTTAGGTGTCATACATGTAAGTGTCTT 4695  
Qy 133 TTTTATATCTCGTATGCCAGTTTGAATTTGAGTATAGAGGTAATCTGATAGAGTTTGAAA 192  
Db 4696 TCTATATCAGTAGTTTATATGAAAACTTTAAGTATAGAGGCTACTTGGATGAATCATTTG 4755  
Qy 193 AAACATTTAACTATAGTATTAATTTTGGCAATTTTTCTTAGCGCAGTATCAATTTTGTGG 252  
Db 4756 CAACTGTCAAATATTTGTTTCATATTTGCTCTAAATTTGCAACATTTCTCTCGTTTTTTCAG 4815  
Qy 253 AGAATAATTTTCGACATTTCAAGACGTTGTCGCGTATTTTACATTAATAAATTCGTTT 312  
Db 4816 ATGGAAGTTTTTCAATCTCAGTCGCGGACTTCTTTACGTCACCTGATTTTCAGGTGTT 4875  
Qy 313 TGTATACCTTATTTAAACGTAATTTAAGCAGTTTTAAGGATAGCTTTCTATTTTTCGCAA 372  
Db 4876 TCTTATACGTTACAAATACCTGTTCTTAAAGTATTTCCGTTTCATCTATTTATACACGTC 4935  
Qy 373 TCTATCAAAAAAGAGCAGTTCTTAATTTACAAAGGCTGAACGATGGGAAAAATATGCAAGTTT 432  
Db 4936 AAAGTAACCAAGAAATATTTCTCTGATTTCTGATCAAGCACGCTCTAGAAAAAT----- 4985  
Qy 433 TATTTGNAATCACATAAACAAATTTCAAAAAAATCTTGTTCATTTGTCATTTGTTAGGTACAG 492  
Db 4986 --GTTTTGCTCTGATGAAGACAATATGATGGTAGGATTACAGCAGTTTGTGCTTTGG 5043  
Qy 493 AAATAGATAAAATTAATTTTATCAATATACCGCTCTTAATTTTCTGTGGGAAGAGCTATAGAGT 552  
Db 5044 ATAATCCTTATTTACGGATCCATTTATCAAGAGTGTAAACCTGAAATTTGATTTGAAT 5103  
Qy 553 TTTCAACAGGGAAGTGGTCGACCAAGCTTTTATAAATCTACCAAGTGAAGTTTTTGAAGC 612  
Db 5104 ATGCGACACACTCAGTAGTAGACCAAGTTTTTGATTAATCTGCCCAAGTGAAGTACAA 5163  
Qy 613 TAAAGCAATTCGTTTCAGATTTTGGATTTGATGTTAGTATTTGATTAAGCGTTGATTAAT 672  
Db 5164 TTTGGGATTAATGCAATCACCTTTTGAAGTTAAGGATTCAGGATTCAGTATCCATTTAATTT 5223  
Qy 673 CATTCGGTTTTACTGCGTTTGAATAAAACAAAAAAATCCAACTGCTAGTGAACCATAGCAT 732  
Db 5224 CACTTGAATTTATAGAGTAAGGTTGAAGAGCGTATCCAAACAATTCGGTCTCTTCAAAGT 5283  
Qy 733 TAACTTTTCCACAAATTTTATAAGCTTAGTCAATCATATGATGAAGAAACGACTTTTGATA 792  
Db 5284 TTACGTTTTTCAACGCAATTTTATAGCTATGAGATATCTTTGGCAAAACGTTTCTCTGATA 5343  
Qy 793 TACTCGAGCGGTAGTGGGCTTAATTTTGTGCTATAGTTTCTATTTTGTAGTTCCAA 852  
Db 5344 TCTGTGAGCCCTAGTTGGTTTGGTCTCTGTGGATTTGGAAATCTTCTTTATCCAC 5403  
Qy 853 TTATTCGTAGAGATGGTGGACCGGCTATTTTCTCAGAAACAGTTTGGACAGAAATCGAC 912  
Db 5404 TTATTCGTAGGATGGTGGGCGAGCCATTTTCTCAAGACCGTGTGGGAGAAATCGAC 5463  
Qy 913 GCATATTATCAATCTACAGTTTTCGATTCGATGATGTTGATGCTGAGGAGCGGCAAAAG 972  
Db 5464 GTATCTTCAAGTTTATAAATTCGTTCTATGTTGTTGATGGGAGAAATCAAGAAGA 5523  
Qy 973 ACTTGCTCAGCCAAACCAAGATGCAAGGTTGGGTATGTTTTTAAATCGGGAACAGATCC 1032  
Db 5524 ATTTGATGGCACAGAATCAAAATGCTGGTGGTATGTTT-----AAGATGGAATGATGCC 5578  
Qy 1033 TAGAATTTACTCCCAATTTGGACATTTTCATACGCAAAAAACAAAGTTTAGACGAGTTACCAAGT 1092  
Db 5579 ACGTATTACCAAAATTTGGACATTTCAATTCG-THAAACGAGCTCTTGATGAATTCACAT 5637  
Qy 1093 TTTATTAATGTTTTAAATTTGGCGATATGAGTCTAGTTGGTCAACGTCACCTTACAGTTGATG 1152  
Db 5638 TTGA-----TGTTAGGTGATATGAGCTTGGTAGGAACACGTCACCAACATTTGATG 5689  
Qy 1153 AATTTGAAAAATATATCTCTCGGTCAAAAGAGACGATTGAGTTTTTAAACCAGGGATTACAG 1212

Db 5690 AGTACGAATCTTATACACCGGAACAAAACGTCGCTCAGCTTTAAACCCAGTATTACTG 5749  
Qy 1213 GTCTCTGCGAGGTAGTGGTCTAGTAATATACACAGACTTCGACGACGTAGTTCGGTTGG 1272  
Db 5750 GTCTTTGCGAAGTAAGCGTGAAGTGAATTAATCTGATTTTGAAGTGTGTAAGACTAG 5809  
Qy 1273 ACTTAGCATACATGATTAATTTGGAATCTCTGTCAGATATTAATAATTTTAAAGACAG 1332  
Db 5810 ACGTGTCTTAATTTGACGAGTGAACATCTGGCGGATATTAAATCTTTATTGAAACAA 5869  
Qy 1333 TGAAGTTGTATTGTTGAGAGGCGGAAGTAAGTAAAGTATATGAAAGTTTGTGTTGTCG 1392  
Db 5870 TAAAGTAGTAGTAATGAAGGATGGACCAAGTATGCGCTTCTCTCCATTTCTTTTAATGG 5929

RESULT 25  
ABA01433  
ID ABA01433 standard; DNA; 12786 BP.  
XX AC ABA01433;  
XX XX  
XX 21-FEB-2002 (first entry)  
XX DE Streptococcus thermophilus eps3 operon #1.  
XX KW Exopolysaccharide; lactic acid bacterium; eps; fermented food product;  
XX KW ds.  
XX OS Streptococcus thermophilus.  
XX XX  
XX WO200179500-A2.  
XX XX  
XX 25-OCT-2001.  
XX XX  
XX 18-APR-2001; 2001WO-FR001199.  
XX XX  
XX 18-APR-2000; 2000FR-00004972.  
XX XX  
XX (INRG ) INRA INST NAT RECH AGRONOMIQUE.  
XX PA (DANO-) CIE DANONE SA GERVAIS.  
XX PA (RHOD ) RHODIA CHIM.  
XX XX  
XX Rallu F, Besancon-Yoshpe I, Frenaux C, Mengaud J, Renault P;  
XX XX  
XX WPI; 2002-017616/02.  
XX XX  
XX New nucleic acid fragments containing exopolysaccharide operon, useful  
XX PT e.g. for increasing exopolysaccharide synthesis in lactic acid bacteria.  
XX XX  
XX Claim 11; Page 39-43; 144pp; French.  
XX XX  
XX The present sequence is an eps operon from Streptococcus thermophilus.  
XX CC Proteins encoded by the eps operon function in exopolysaccharide (EPS)  
XX CC synthesis. The operon is useful for producing chimeric eps operons, for  
XX CC optimising production of EPS in lactic acid bacteria. EPS impart texture,  
XX CC mouth feel and rheological properties to fermented food products (e.g.  
XX CC yoghurt). They function as thickeners, to provide free-flowing and creamy  
XX CC texture, and may also have biological activities beneficial to health  
XX XX  
SQ Sequence 12786 BP; 4161 A; 2228 C; 2047 G; 4055 T; 0 U; 295 Other;

Query Match 4.9%; Score 346; DB 6; Length 12786;  
Best Local Similarity 57.2%; Pred. No. 1.1e-37;  
Matches 727; Conservative 0; Mismatches 529; Indels 16; Gaps 5;

Qy 123 TATTTTGCATTTTATATCTCGTATGCCAGTTGAATTTAGTATAGAGGTAATCTGATA 182  
Db 4649 TANGTGTCTTACTATATCAGTAGTTATTATGAARATCTTAAGTATAGAGGCTACTGGAT 4708  
Qy 183 GAGTTTGAAGAAACATTTAACTATAGTATAATATTGCAATTTTCTTACGCGCAGTATCA 242  
Db 4709 GAACTCATTCGCACTGTCAAATATTGTTTCATATTGCTCTTAATTGCAACATTTCTCTCG 4768

Qy 243 TTTTGTGGAGAAATAATTTCCGCACTTTCAAGACGTGGTGCCTGTATTTTCCATTAATA 302  
Db 4769 TTTTTCGAGATGGAAGTTTTTCAATCTCACGTGCGGACTTCTTACGTCCACCATGATT 4828  
Qy 303 AACTTCGTTTGGTATACCTATTTAAAGTAATTAATTAAGCAGTTTAAAGTAGCTTTCTA 362  
Db 4829 TCAGGTGTTCTCTTATACGTTTACAAATACTGTTCTTAAGTATTTCCGCTCATTTAT 4888  
Qy 363 TTTTCGCAATCTATCAAAAAAGACGATCTTAATTAACAACGGCTGAACGATGGAAAT 422  
Db 4889 ACAGCTGTAAGTAACAGAAATATTTCTCTTGATTTCTGATCAGGACGCTCTTGATAT 4948  
Qy 423 ATGCAAGTTTATTTGAATCAATAAACAATTAACAAAAATCTTTGTCATTTGGTAGTT 482  
Db 4949 GTTNTTATCTCGTATGAAGGACAATATGATGGTAGGA-----TTTCAGCAGTTTGT 5000  
Qy 483 TTAGGTACAGAAATAGATAAAATTAATTTATCATACGCTCTATTTCTGTGGAAGAA 542  
Db 5001 GTCTTGGATAATCTCTTATTTTCACTGATCCATTTTATCAAGAGTGTAAACCTGAAAATTTN 5060  
Qy 543 GCTATAGAGTTTTCACAAGGGGAAGTGTGCACACGCTCTTTATAAATC-TACCAGTGA 601  
Db 5061 GATTGAATATGCCGACACACTCNAGTAGTAGACCAAGTTTGTATTAATCNTGCCAAGTGA 5120  
Qy 602 GTTTTATAGCGTAAGCAATTCGTTTCAGATTTTGGAGTTGTAGGTATTTGATGAAGCGT 661  
Db 5121 GCAGTACAAGATTTGGGATTATGCGTCACCATTTTGAACCTTATGGGATCCAGTATCCAT 5180  
Qy 662 TGATATTAATTCATTCGGTTTTTA-CTGGTTGAAAAACAACAAAAATCCAACTGCTAGGTG 720  
Db 5181 TAATTTGAATGCCCTTGAAATTTATGAGTCAGGGTGAAAAACAATATCCAAATTTGGGTC 5240  
Qy 721 ACCATAGCATTTGTAATCTTTTCCACAAATTTTATAAGCCTAGTCATATCATGATGAAC 780  
Db 5241 CTTTCAAGTTGTACGTTTCTAGGCATTTTATAGCTATGGAGATATCTTGGCGAAAC 5300  
Qy 781 GACTTTTGGATATATCTCGGAGCGGTAGTCGGGTTTAAATTTATTTGGTATAGTTCTATTT 840  
Db 5301 GCTTCTCGATATCTGTGGAGCTCTAGTTGGTTGGTGTCTGTGGTATTTAGGAATCT 5360  
Qy 841 TGTAGTTCCAAATTTATTCGTAGAGATGGTGACCGGCTATTTTTCCTCAGAAAAAGAGTTG 900  
Db 5361 TCCTTTATCTCTTATTCGTAAGATGGGGACCAAGCTATTTTTCCTCAATACCGTGTGG 5420  
Qy 901 GACAGATGGACGATATTTACATTTACAGTTTTCGATCCGATGATGCTGCTGATCTGAGG 960  
Db 5421 GTGAAATGGTCGTATCTTTAAGTTTATATAAATTCGGTTCTATGCGTGTGATGCGGAAG 5480  
Qy 961 AGCGCAAAAAAGACTTGTCTCAGCCAAAAACAGATGCAAGGGTGGGTATGTTTAAAAATGG 1020  
Db 5481 AATCAAGNAGAAATTTGATGGCAGAGATCAATGTCTGGTGTATGTTT-----AAGAT 5535  
Qy 1021 GAAAAACGATCTTAGAATTTACTCCAAATTTGGACATTTTCATACGCAAAAAACAAGTTTAGACG 1080  
Db 5536 GGACAATGATCCAGTATTTACCAAAATTTGGACATTTTCATTCTG-TAAAAACAAGTCTTGATG 5594  
Qy 1081 AGTTACACAGTTTATTAATGTTTAAATTTGGGATATAGTCTAGTTGGTACACGTCAC 1140  
Db 5595 AACTTCCACAATTTTGGAAATGTTCTTAAAGGTGATATGAGCTTGGTATGGGACACGTCCTC 5654  
Qy 1141 CTACAGTTGATGAATTTGAAAAATATATCTCTCGTCAAAAAGAGACGATGAGTTAGTTTAAAC 1200  
Db 5655 CAACAGTTGATGATATGAAAAATATACACTGAACAGAAACGTCGTTAAGTTTAAAC 5714  
Qy 1201 CAGGATTTACAGGTCTCTGGCAGGTTAGTGGTCTAGTAAATATATACAGACTTCGACGACG 1260  
Db 5715 CTGGTATCACAGGTCTTTGGCAAGTAAGTGGACGAAGTGAATCACTGACTTTGATGAGG 5774  
Qy 1261 TAGTTCGGTTGGACCTTAGCATAACATTTGAATTTGGACATATCTGTCAGATATTAATAATTT 1320  
Db 5775 TTGTTAAACTGGATGTTGCTTATATTGATGGTGGCAAACTCGGCGAGATATTCAAATCT 5834

Qy	1321	TATTAAAGACAGTGAAAGTTGTTATTTGTTGAGAGAGGGAAGTAAGTAAAGTATATGAAAG	1380
Db	5835	TACTGAAACCGATTAAAGTAGTAGTAATGAAGGATGGAGCAAGTGATGGCTTTCACCAT	5894
Qy	1381	TTTGTTTGGTCG	1392
Db	5895	TTCTTTTAATGG	5906

Search completed: April 29, 2005, 02:36:47  
Job time : 3285 secs







```

; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 2303:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1443 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1443
; SEQUENCE DESCRIPTION: SEQ ID NO: 2303:
US-09-107-433-2303

Query Match 5.9%; Score 415.8; DB 4; Length 1443;
Best Local Similarity 58.3%; Pred. No. 5.5e-66;
Matches 789; Conservative 0; Mismatches 552; Indels 12; Gaps 3;

Qy 14 ATTGGCAATATTGATATGATAGACAGTTGCAATTTCTGCAATCTTAACAAGTCATATACC 73
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
102 ATTGGCAATGTCAGAGTTTCTTGTATTTTATTGCGCTTATCTACTTAGCACTGTGAG 161
Qy 74 AAATGCTGATTTAAATCGTCTCGAATTTTATCATATGATGTTTCAATTTATTTGCAAT 133
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
162 AGAACAGAGATTGTTTCAACAACAGCTATGTTACTTTATCTTATCTCCACTATTTTGCCTT 221
Qy 134 TTTTATATCTCGTATGCCAGTTGAAATTTGAGTATAGAGGTAATCTGATAGAGTTTGAATA 193
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
222 TTATATCAGTATTATGACAGAGATTCTTTAAAGGGGATATTTGATTGAATCTGTCCA 281
Qy 194 AACATTTAATATAGTATATATTTGCAATTTTCTTACGGCAGTATCATTTTGTGGA 253
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
282 GACATTTGAATATATCTCTTCTTTGGGCTAGCGATTAGTATTTCTAATTTTCTTTAGA 341
Qy 254 GAATAATTTCCGACCTTTTCAGACGCTGTCGCTGTTTTCACATTAATAACTTCGTTT 313
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
342 AGATCGAATTAGTATTCCAGAAGAGGATGATTACTTCTCTCTATATACATGTTCTCTT 401
Qy 314 GGTAATACCTATTAAACGTAATTTAAGCAGTTTAAAGGATAGCTTTCTATTTTCGCAAT 373
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
402 AGTCTATGTCCTAAACCGATTATCAAGTGTATGGAAGGGCTTATCCCACTTTAA 461
Qy 374 CTATCAAAAAAGACGATTCTTAATTAACAACGGCTGAAACGATGGGAAAATATGCAAGTTT 433
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
462 AGGAAGTAAGAAGATTCTCTACTTACAGCAACTCTCGTGTGCAAAAGGATTGGATAG 521
Qy 434 ATTTGAATCATATAACAATTTCAAAAAATCTTTGTCATTGTTAGTTTAGTACAGA 493
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
522 ACTAATGAATCAGATGTTGTTGGGAGTTGGTAGCCGTCAGTGTTTTAGATAAAC 581
Qy 494 AATAGATAAAATTAATTTATCATACCGCTCTATTTATTTCTGTGGAAGAAGCTATAGAGTT 553
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
582 AGATTTTTCAGCATGATTATTAAAGGTTGTAG-----CAGAGGGGAGATCGTAACCTT 635

```

```

Qy 554 TTCAACAAGGAGTGGTGCACACCGCTCTTTATAAATCTACCAAGTGAGTTTTTAGACGT 613
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
636 TCGCACTCATGAGTGGTGCATGAAGTCTTTATCAATCTTCCAAAGTGAATAATACAAATAT 695
Qy 614 AAAGCAATTCGTTTCAGATTTTTCAGTGTGTTAGTATTGATGTAAGCGTTGATATTAATTC 673
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
696 TGGAGAGCTTGTCTCTCAGTTTGAACAGATGGAGTTGATTAACAGTCAATCTTAATGC 755
Qy 674 ATTCGGTTTACTGCGTTGAAAAACAAAAATCCAACCTGCTAGGTGACCATAGCAATGT 733
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
756 TTTTCGATTGTATCTTGGCACATAAACAAGCAAAATTTGTGAGATGGCAGGACTAAACGTTGT 815
Qy 734 AACTTTTCCACAAATTTTATAAGCCTAGTCAATATCATGATGAACAGCACTTTTGGATAT 793
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
816 GACTTTTCTACAACAATTTTATAAGACTAGCCATGTGCTTAAGCGGGTTATTGATAT 875
Qy 794 ACTCGGAGCGGTAGTCGGGTTAATTTTGTGTGTATAGTTTCTATTTTTGTAGTTCCAAT 853
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
876 TATCGGTTCCCTGTTAGGTTTGATATCTATGTGTGTCTAGTCAGTATTGTACTGGTTCCTT 935
Qy 854 TATTCGTAGAGATGGTGGACCGGCTATTTTGTCTCAGAAACGAGTTTGGACAGATGGACG 913
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
936 GATTCGAAAGGATGGGGCTCTGCTATTTTGTCTCAGACGCGTATTGGGAAAAATGGTCG 995
Qy 914 CATATTTACATCTACAAGTTTCGATCGATGATGTTGATGCTGAGGAGCGCAAAAAAGA 973
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
996 CCATTTCACTTTTACAAGTTTCGCTCTATGTGTGTGATGCTGAGGAGAAAAAAGAGA 1055
Qy 974 CTTGCTCAGCAAAACACAGATCAAGGGTGGGTATGTTTAAATGGAAAAACGATCCT 1033
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1056 ACTCATGGAAACAAATATACCATGCA-----GGGTGGAATGTTTAAGGTGATGAGATCCA 1110
Qy 1034 AGAATTAATCCAATTCGACATTTTATACGCAAAACAAAGTTTATAGACAGTTTACCACAGTT 1093
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1111 CGTATCACGAAATTTGTCATTTTATACG- GAAGACGAGCTTGGACGAGCTACCACAGTT 1169
Qy 1094 TTATAATGTTTAAATTTGGCGATATGATCTAGTTGTTGATACAGTCCACCTACAGTTGATGA 1153
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1170 TTACAATGTTTCAAGGGAGATATGATTTGGTAGGACACGACCAACCAAGTGGACGA 1229
Qy 1154 ATTTGAAAAATATACCTCTCTGTCAAAAGAGACCATTTAGTTTAAACACAGGGATTACAGG 1213
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1230 GTATGAGCACTATACCCGACAAACAAACGTCGCTCAAGTTTAAACCTGGCATACAGG 1289
Qy 1214 TCTCTGGCAGGTTAGTGGTGTGTTAGTAAATATACAGACTTCGACGAGTGTTCGGTTGGA 1273
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1290 TCTATGGCAGGTGAGCGGAGGAGATCAAGAAATTTTCGATGAGGTTGTCAAATTTAGA 1349
Qy 1274 CTTAGCATATGATTAATTTGGACTATCTGCTCAGATATTAAATTTTATTAAGACAGT 1333
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1350 TGTGGCCTATATAGACGGTTGGCAATTTGGAAAGATATTGAAATTTTATTGAAGACAGT 1409
Qy 1334 GAAAGTTGTATTCTTCAGAGAGGGAAGTAAGTA 1366
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1410 TAAAGTTGTATTGATGAGGATGGAGCGGAAGTA 1442

```

```

RESULT 4
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA

```





Db 924581 TTGAAATTAGCAAAATGGCTGTGAATCTTAAAGAGTGTAGTAGTAATGCTTAAGACA 924640  
Qy 983 ---CCAAACACAGATGCAAGGGTGGGTATGTTTAAATGGGAAAAACGATCCTAGAAAT 1039  
Db 924641 GATCCTGAAGCAAGAGCTGAATGGGAAAAAGATTTTAAATTAAGAATGATCCTAGAAAT 924700  
Qy 1040 ACTCAATTGGACATTTTCATACGCAAAACAGTTTATAGACAGGTATACACAGTTTATAA 1099  
Db 924701 ACTAAATCGGTGCTTTTATTCGC-AAACAAAGTTTAGATGAATTACCTCAATTTATTTAA 924759  
Qy 1100 TGTTTAAATGGCGATATGATGTAGTTGGTACAGTCCACCTACAGTTCGATGAATTTGA 1159  
Db 924760 TGTTTTAAAGGTGAATGAGTTTGTGGCTCGCCCAATGTAATGATGAGCTAGA 924819  
Qy 1160 AAAATATACCTCTGGTCAAAAGAGACGATGAGTTTAAACACAGGGATTAACAGTCTCTG 1219  
Db 924820 GCGTTATGAAGAAACGTAGATTATTACTTAATGGCAAGACCCGGTATGACAGGCTTTG 924879  
Qy 1220 GCAGTTAGTGGTGTAGTATATACAGACTTCGACGAGTAGTTCGGTTGAGCTTAGC 1279  
Db 924880 GCAAGTCAGCGGAGCTAATAATATC---GATTATATATCTCGCGTTTATTTTGTGATCTTG 924936  
Qy 1280 ATACATTGATAATTGGGACTATCTCGTCAGATATTAAAAATTTTATAAAGACAGTGAAGT 1339  
Db 924937 GTATGTAATAAATTTGGTCACCTTGGAAATGATTTGCGATTTTATTTTAAACAATGAATGT 924996  
Qy 1340 TGTATCTGTGAGAGAGGGAAGTAAGTAAAGATATAT 1375  
Db 924997 GGTTTTAAATCGTATGAGGCTTATTAAACGTTATTT 925032

RESULT 6

US-09-107-532A-2527  
; Sequence 2527, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 2527:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 984 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double

; TOPOLOGY: circular  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (B) LOCATION 1...984  
; SEQUENCE DESCRIPTION: SEQ ID NO: 2527:  
US-09-107-532A-2527  
Query Match 2.2%; Score 152.6; DB 4; Length 984;  
Best Local Similarity 51.4%; Pred. No. 1.6e-18;  
Matches 384; Conservative 0; Mismatches 354; Indels 9; Gaps 1;  
Qy 4191 AAAAAATGAATGATTTAAATTCAGTTATTGTACCAATTTTAAATGTCCAGATTATCTTGA 4250  
Db 12 AAATATAATGTGTGAGATTAGTATTATTTCTGTGTATAAAGTTGAAAAATATTTAAG 71  
Qy 4251 TAAATGTATTAAACAGTATTATTAAACCAACATATATCTAAATTTAGAGTTATCTCTGAAA 4310  
Db 72 AAAATGTAGATTTCGATTTTAGCCCAACATTTACTGATTTTGAAGTTATATATTAGTAGA 131  
Qy 4311 TGATGGAAGTACTGATGATTTCTGAGAAAATTTGCTTAAACTATATGAAGAACGATGGAAG 4370  
Db 132 TGATGATCTCTCGACAAATAGCGGGAAAAATCTGTGATGATGATGCTGAAAAAGATATCG 191  
Qy 4371 AATTAAATATTACAAGAAAAATTAATGCGGCTCTAGCAGATGCTCGAAAATTTCCGACTAGA 4430  
Db 192 AGTACGTGTATCCATTAAGAAAACCGCGGATTAAGCAGTCTGAGAAATCCGGGGATTGA 251  
Qy 4431 ACATGCAACAGGTAATAATATTGCTTTTGTTCGATTTCTGATGACTATATAGAAGTTGCAAT 4490  
Db 252 TGTTCGAGAGGCAAAATCTTAGGCTTTTGTGACAGTGTATGATATATATTCGATGAAGATAT 311  
Qy 4491 GTTCGAGAGATGTCATGATAATACTAGTATTAATGCGGATATAGCAGATAGATAGATTT 4550  
Db 312 GTATGAATTTTGTATGAGAAATTTGAAAATTTCTATGATGCGGACATTAGCTCTGTTGAAT 371  
Qy 4551 TTCTTTAGTAGACGAAAAACGGGTATACAAAGAAAAAGAAATAGTAATTTTCATGTCCTT 4610  
Db 372 GATTCCTTTTATGGTGTAGATATTAAGAAAGCTTAATAAGAAAAGAAAGTAAATATTAT 431  
Qy 4611 AACGACAGAGAGACTGTAAAGAAATTTTGTCCAGATCTTAATATAGAAAATATGTTTG 4670  
Db 432 AAATAAAAAGGAAGCAATAAAATCTGTTTAGAAGGAACGCAATTTTATGCATACGCATG 491  
Qy 4671 GTGCAAGCTTTTATTCAGGAGATATTATAAAGATATAAATTTCCAAATTAATATAGAAG 4730  
Db 492 GAATAAACTTTATCGTAAGAACTTTTAAAGATATATCGTTACTTGGATGGAAAAACATT 551  
Qy 4731 TATTGCTGAGGATTTGCTTTTAAATTTGGAGGCTTTTGAACAATGTAAACAGCTGTAGTAGT 4790  
Db 552 TGAGGATGCATATATAATATAGATCTATTATT-----TCAACACAGAAAAATGTT 602  
Qy 4791 TGATACCTAGAGAAATATTATTATAATTTATGTCATTTCGTAACAGTTTCGCTTATTAATCAGAA 4850  
Db 603 TGTTTCCAATGAAGAGAAATATTCTTCTTCAAGAAATGATAGTATTATTGGAAGAAAG 662  
Qy 4851 ATTCTCTATAAATAATATTGATTTAGTCACAGATTTAGGAGAAATACCCCTTTAAGTTAAA 4910  
Db 663 TTTTCTTATGAATAATTTTCGATGTTTATGAAGCATGCGAATATAACAAAGAGAAGATTTT 722  
Qy 4911 AAGAGAGTTTAGTCATTATTTTTCATGTC 4937  
Db 723 AGATGCTTTTCCAGATTTCATGATTC 749

RESULT 7

US-09-134-000C-2987  
; Sequence 2987, Application US/09134000C  
; Patent No. 6617156

## ; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; FILE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 032796-032  
; CURRENT APPLICATION NUMBER: US/09/134,000C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/055,778  
; PRIOR FILING DATE: 1997-08-15  
; NUMBER OF SEQ ID NOS: 6812  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2987  
; LENGTH: 993  
; TYPE: DNA  
; ORGANISM: Enterococcus faecalis  
US-09-134-000C-2987

Query Match 2.1%; Score 145.4; DB 4; Length 993;  
Best Local Similarity 52.0%; Pred. No. 3.2e-17;  
Matches 382; Conservative 0; Mismatches 341; Indels 12; Gaps 2;  
Qy 6164 TATTAGTAAATTTCTATAATTTGACCTATATATATAGTAAATATTTATCTAAATG 6223  
Db |||||  
Qy 18 TATGCCAAATTTAGTATTTTCTCTATACATGTAGAAATATTTAGAAATG 77  
Db |||||  
Qy 6224 TATAGATAGCATTTGTAATTCAGACTACAAACATATAGAGATTTCTCTGGTGAATGACGG 6283  
Db |||||  
Qy 78 TGTACGCTCTATTTAGTCAACGTTTACTGACTTTGCAATTAATTTCTGGTGGACGATGG 137  
Db |||||  
Qy 6284 TAGTACGATTAATTCGGAAGAAATTTGTTAGCATATCGGAAGAAAGATAGTCGATTCG 6343  
Db |||||  
Qy 138 CTCTCCAGACAGTTCTCGGACAAATGTGTGATCAGTTTCTGCAACAGATCAACGGTAAA 197  
Db |||||  
Qy 6344 TTATTTTAAAAAGAGAACGGCGCTATCAGATCGCCGTAATATATGATCAATAGTCGGC 6403  
Db |||||  
Qy 198 AGTTATCATTAAGAAATGTGGCTAAGCATGCTGTAATGCTGGAATGAAATAGC 257  
Db |||||  
Qy 6404 CAAGGGTGACTCTTAGCTTTTATAGACTCAGATGATTTTATTCATTCGGAGTTTCATCCA 6463  
Db |||||  
Qy 258 AACAGGTGAGTATTAGGTTTCGTAGATGATGATTTACATTCGAGATGATATGATGA 317  
Db |||||  
Qy 6464 ACGTTTACAGAAATGTAGAGAGAAATGCGCTTGTGGCAGTTGCTGTTGATGATAG 6523  
Db |||||  
Qy 318 ACTATTATATACAAATATCGTAAAGAAAGATGCGGATTTATCAATCTGTGTTATTTATGA 377  
Db |||||  
Qy 6524 GSTAGATCTTCGGGCAATTTCTTAACAGAGAGCGCTTCTCAAAATCAGGCTGTCT 6583  
Db |||||  
Qy 378 TGTCTATG----AGGGGAAAGAGCCAAATGTAAAGATTTAATACAGGGAACCTTTTCCA 433  
Db |||||  
Qy 6584 GAGCGGCAAGGAATGTTTGTAAAAAGCTGCTAGAGCGCGATGGTCTATCGCTTTGTGGTGGC 6643  
Db |||||  
Qy 434 GAGAAGAGCAATATTGTTAA-----TTTTACAGGAATATTATCTCGTACATGC 485  
Db |||||  
Qy 6644 CTGTAATAAATCTATAAAAAAGAACTATTGTAAGATTTTCGATTTGAAAGGGTAAGAT 6703  
Db |||||  
Qy 486 TGTAAATAGTTATATAAACGAAACTCTTTGCGGATTTAAGATATCTAAAGGAAAGTA 545  
Db |||||  
Qy 6704 TCATGAAGATGAATACITTCATCTATCGCTTCTCTATGAGTTAGAAAAGTTGCAATAGT 6763  
Db |||||  
Qy 546 TCATGAAGATTCATTTATATATAGTAGATTTGTTGTCGAATGTGCAAGATCTCTATTGA 605  
Db |||||  
Qy 6764 TAAGGAGTGTGTTGACTATTATGTTGACCGAGAAAAATAGTATCAAACTTCTAGCATGAC 6823  
Db |||||  
Qy 606 TTCAACACAAAAATATTACTATTATCATCGAATGGGCGAGCAATTAATACAGAAACATTTTC 665  
Db |||||  
Qy 6824 TGACCATCGCTTCCATTCGCTACTCGGAATTTCAAAATGAACGAATGGACTTCTATGAAG 6883  
Db |||||  
Qy 666 AGATAAACAAATTTGAGTTTATAGAGCTTTGGAGAGAAATGAACTTAAATTTGAAGGTAA 725  
Db |||||  
Qy 6884 TAGAGGAGATAAAGA 6898  
Db |||||  
Qy 726 GGGAGCAGTTTATAGA 740  
Db |||||

## RESULT 8

US-08-961-527-71  
; Sequence 71, Application US/08961527  
; Patent No. 6420135  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 391  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961,527  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB340P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 71:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 32768 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-961-527-71

Query Match 2.1%; Score 144.6; DB 3; Length 32768;  
Best Local Similarity 52.3%; Pred. No. 9.1e-17;  
Matches 368; Conservative 0; Mismatches 329; Indels 6; Gaps 2;  
Qy 5144 AGAAATTTCAAACGACGTAGAGGTAAATAATGATAAATAATTAGTGTATTGTTCCAGTTTA 5203  
Db |||||  
Qy 6153 ACATAAATCAAGGAGAAAAATGCTAGTGGATGATAAATAACAGTCATTGTACCAGTATA 6212  
Db |||||  
Qy 5204 TAATGTAGATAAATATTTAAGTAGTTGTATAGAAAGCAATTAATTAATCAAAATTTAAAAA 5263  
Db |||||  
Qy 6213 CAATGTGGAACACTATCTGAGGAAGTGCCTAGATAGTATTATTACTCAACATATAAAAA 6272  
Db |||||  
Qy 5264 TATAGAAATATTTATGATAGATGATGCTCTGTAGATGATTCTGCTAAAAATATGCAAGCA 5323  
Db |||||  
Qy 6273 TATTGAGATTTGTCGTGTTAATGATGTTCTACCGATGCTTCAGGTGAAATTTGTAAGA 6332  
Db |||||  
Qy 5324 ATATCGAGAAAAAGATAAAGAGTAAATAATTTTTTTCATCAATCATAGTGAGATATCAAA 5383  
Db |||||  
Qy 6333 ATTTTCAGAAATGATCACCAGAAATCTCTATATAGAACAGAAAAATGCTGCTTTCTGTCG 6392  
Db |||||  
Qy 5384 TGCTAGAAATCATGGAATTAACGGAGTACAGCTGAATATATATTATTTGTTGTTGACTCTGA 5443  
Db |||||  
Qy 6393 CGCAGAAACACCGCTCTGAAATATATATGTCGGAAATTTATGTGACCTTTGTGGACTCGGA 6452  
Db |||||  
Qy 5444 TGATGTTGTCATAGTAGATTAGTAGAAAAATTTATATTTTAAATATATAAAGTAGAAG 5503  
Db |||||  
Qy 6453 TGATTGGATTGAGCAAGATTATGTAGAACTCTATATAAANAATATAGTAGATTCAGGC 6512  
Db |||||  
Qy 5504 TGATTATCTGTTGTTGTTGACGCTACTTTTTCAGAAAAATATAAATAATTTTGAAGTCAA 5563  
Db |||||

Db	6513	TGATATTGCAGTTGGTAATTTATTTCTTTCAACGAAGTGAAGGAATGTTCTACTTTCA	6572
Qy	5564	TA-ATCCAAATATTGATTTTTGAAGCAATTAATACCGTCGAGCATGGGAGAAAAAATT	5622
Db	6573	TATATTGGGAGACTCCCTATTATGAGAAGTATATGATAATGTTTCTATCTTTGAGAACTT	6632
Qy	5623	TTATGAATTTGTATATATAAATAATATTTTTCTA-----CTCCCTGTTTGTAAACTATATAA	5677
Db	6633	GTATGAAACTCAAGAAATGAAGAGTTTTGCTTTGATATCTGTTGGGGTAAACTCTATAA	6692
Qy	5678	GAAAGATACATAACAGATCTTTTTTCAAGAGAATCAATGGTTTAGGGAAGATTTACTTTT	5737
Db	6693	GGCAAGATTCTTTTGAGCAGTTGCGCTTTGCATAGTAGGTAAATTAGGAAGATGGTTACCT	6752
Qy	5738	TAATCTGCATTATTAAAGAATATAGATAGAGTTTAGTTATTATTCGCTGAACATCTTTATTTT	5797
Db	6753	CAATCAAAAGGTATATTTATTATCAGAAAAGGTATTTTATTTTAAATAAAAGTCTTTATGC	6812
Qy	5798	TTATAGGAGAGGTATACTAAGTACAGTAAATCTTTTTAAAGAA	5840
Db	6813	TTATCGGATTAGAAAAGGTAGTTTATCAAGAGTTTGGACAGAA	6855

## RESULT 9

```

US-09-134-000C-2344
; Sequence 2344, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2344
; LENGTH: 2529
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-2344

```

Query Match	2.0%;	Score 139.2;	DB 4;	Length 2529;
Best Local Similarity	44.7%;	Pred. No. 5.1e-16;		
Matches 677;	Conservative 0;	Mismatches 823;	Indels 14;	Gaps 3;
Qy	4198	AATGATTTAATTTTCAGTTATTGTACCAAAATTTATAATGTCOAAGATTAATCTTGATAAATGT	4257	
Db	25	AATGTGAAGTGTGAGTCATTGTACCGCTCTATAACGTAGAAACCTATTTAGGAAGGCCA	84	
Qy	4258	ATTTAACAGTATTATTAAACCAAAATATACTAAATTTAGAGGTTATTTCTCGTAATATGATGGA	4317	
Db	85	TTAATGAGTTTGAATAAACCAAAATTCGAAAGAAATCGAAATTTCTTAATTAATTAATGATGCC	144	
Qy	4318	AGTACTGATGATTCCTGAGAAAATTTGCTTAAACTATATGAAGACGATGGAAGAAATTTAAA	4377	
Db	145	TCTACTGATAATAGTCAAAAAATTTATTGAAGAAATTCACAAGATGATCCACGTTTTTCGC	204	
Qy	4378	TATTACAAGAAAATTAATGCGGCTCTAGCAGATGCTCGAAATTTTCGACCTAGAACATGCA	4437	
Db	205	GTTTTTCATGTTTAAAAATGGGGGAATTTGAAAAGCCCTTAACTTGGGTGTTTCAGAAGCA	264	
Qy	4438	ACAGGTAATAATATTGCTTTTGTGCAATCTCGATGACTATATAGAAAGTTGCAATGCTTCGAG	4497	
Db	265	AAAGGTGAGTATATTGCTGAAATTTGAAAGTGACGATTATGTAGCATTTGCATGCGCTATGAG	324	
Qy	4498	AGAAATCATGATAATATAAATCTGAGTATAATGCGCATATAGCAGAGATAGATTTTTTGTTTA	4557	
Db	325	CGTTTTATCAATACGGCTAAAGTCATCATGCAGATGTTGTCCGGTGTAAATTTGGGTTGAA	384	
Qy	4558	GTACACGAAAAACGGGTATACAAAGAAAAAAGAAATAGTAAATTTTTCATGCTCTAAACGAGA	4617	

Db 1451 ATGATAAATATTTTATTAGATACATCAAGAAATAAATCTGGTAAACTTTAGAACAAAT 1510  
Qy 5698 TTTTCAAGAGAAAT 5711  
Db 1511 TTGTTGGCGAGGAT 1524

RESULT 10  
US-09-107-532A-1469  
; Sequence 1469, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD/ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 1469:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1002 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: circular  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...1002  
; SEQUENCE DESCRIPTION: SEQ ID NO: 1469:  
US-09-107-532A-1469

Query Match 2.0%; Score 137.4; DB 4; Length 1002;  
Best Local Similarity 48.0%; Pred. No. 9e-16;  
Matches 480; Conservative 0; Mismatches 496; Indels 25; Gaps 2;

Qy 5153 AAAGCAGTAGAGGTAATAAATGGATAAATAGTGTATTTGTCAGTTTATATAGTAGA 5212  
Db 9 AAAGCAATTGAAACATAAATGTCAGATAAGTATAATAGTTCAGTTTATATAGTGA 68  
Qy 5213 TAAATATTAGTACTTCTATAGAACGATTTAATCAAAATTATAAATATAGAAAT 5272  
Db 69 AAAATATTAAAAAGGTGTGTGATTTCAATTTCTAGGCCAAACATTTACCGAATTTGAAT 128  
Qy 5273 ATTATTGATAGATGGCTCTGTAGATGATTTCTGCTAAAAATATGCAAGGAATATGCAGA 5332

Db 129 ATTATTAATTGATGATGTTCTTACAGATAAATTCAGGAAGTATATCGGATGAACCTTCGAA 188  
Qy 5333 AAAAGATAAAAGAGTAAATAATTTTTTTCACATAATCATAGTCGAGTATCAAAATGCTAGAAA 5392  
Db 189 AACAGACACCGAATAAAGTGATACATAAGGAATAATGGGGGTCTAAGTGATGCCAGAAA 248  
Qy 5393 TCATGGAATAAAGCGGAGTACAGCTGAATATATTATGTTTGTGCTGCTGATGATGTTGT 5452  
Db 249 TATTGGTATTGAGGTAGCAAGCGGAGATTTTATTGGGTTTATAGATAGTGTGATTTACAT 308  
Qy 5453 TGATAGTAGATTAGTAGAATAATTTATTTTAAATATATAAAGTAGAAGTGAATTTATC 5512  
Db 309 TGATGAAGATATGATGCAATTTTATATAACAATATATTGAAATATGATGAGAGTTATC 368  
Qy 5513 TGGTGTGTTGACGCTACTTTTTCAGAAATAATAAATAATTTTGAAGTGAATAATCCAAA 5572  
Db 369 AATGTT-----GGTATTTATGACGCTCCATANGAATAAAGAAT 407  
Qy 5573 TATTGATTTTGAAGCAATAATACCGTCGACGACATGGGAGAGAAAAAATTTTATGAATTT 5632  
Db 408 TAAAAAGTTAACTCTCTTTTCTCAATTGGTTACTAAATCTGAAGCAATTTGAATTAGTATT 467  
Qy 5633 GTATATAAATAATATTTTTTCTACTCTCTGTTGTAACACTATATAAGAAAAATACATAC 5692  
Db 468 AGATGGTAAAGCTAGTTGTTGCTAATGCAAGTAAGCAAGCTATATAAAGAAATTTATTGA 527  
Qy 5693 AGATCTTTTTCAGAGAGATCAATGTTAGGAGAGAGATTTTCTTTTAACTGCTATTATT 5752  
Db 528 GAATGAGATATCCAAAGGAAAAATTTGTAAGACGACGACGATATATCTGAAAAATTTAT 587  
Qy 5753 AAAGAATATAGATAGAGTTAGTTATTTGACTGACATCTTTTATTTTATAGGAGAGTAT 5812  
Db 588 CAATCAGTGTAAATAAATCCATGTTGATCTCTCAAAAAATATTATTATTATCTATCGTGG 647  
Qy 5813 ACTAAGTACAGTAAATCTTTTAAAGAGGTTGTTTTCGAATTTGGAATTTGCAAAA 5872  
Db 648 AGATAGTATTACTAGTAAAAAATTTACAAAAAAGATTTTGTATATAATTTGATGATGGA 707  
Qy 5873 ACAAGTATAGTATTGTTTAAAGCAAAATATATGTTGAGGATTTTTCAGCTATCAATTTGTTAA 5932  
Db 708 TGAGAAATGAAAAATGGATTTATAAATAATTCCTCAATATGACGATAAAGTACATAGTAG 767  
Qy 5933 AGATACTATACGTTGGCAAGTATTATTATATAGCTTACTAATGTTTAAATACGAAAAACA 5992  
Db 768 AGTTGTTGGCTT---ACTTCTGTTGTAGATAAATAGTTTAACTGAAGAAAAAG 823  
Qy 5993 GTCATTTTGGACAAATTTTAAATTTTGAATAATCTTTTATAAATAATATTATTTAACTT 6052  
Db 824 GACTGTATCAAGAAAAAACAAGTAAGAGAGTATTTAATTTGATAAATTTCTTTTATTATA 883  
Qy 6053 GTTAAAGTATCTAACAAAAATTTCTTGTCTAAAAATTTTGTATAGAAATTTGTTCCAA 6112  
Db 884 TAAAAAATAAACGATTAACCTTTTCAAGAAAAATATCTATGTTATTTGTTCTTTTGAA 943  
Qy 6113 CAAAGTTTTTAAAAAATAATTTATGTTTATATAAGAGATA 6153  
Db 944 TTGCTGCTATAAATAACCGTTAAATATCATGATATAGAAA 984

RESULT 11

US-09-949-016-14164/c  
; Sequence 14164, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20

```

; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14164
; LENGTH: 18773
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14164

```

Query Match	1.9%	Score 132.6	DB 4	Length 18773
Best Local Similarity	47.2%	Pred. No. 1.2e-14		
Matches 604	Conservative 0	Mismatches 659	Indels 16	Gaps 6
Qy	2850	AAGATATATAAAGTGGGAAAATCAACATCTTCTCTATTTGCTTTACAGATGATTTCTAT	2909	
Db	18379	AATACATATATATAAATTTATATATATAATAAATTCCTATATAAAATATATATATATATAT	18320	
Qy	2910	TGATGAATTTGCTAAATGCGAGAAATTTAGGTTTTTTTATTTGCTAGAGAGTTAAAAATAGA	2969	
Db	18319	ATATAATATCTATATAAATA-----TATATGATATTATATATAATATTTCTATATA	18268	
Qy	2970	AAATAAATCTAAATTTAAAGAAATTAATTACTAAAAATAAAATAGTTGATTTTGTGAGAG	3029	
Db	18267	AAATATATGTATTTATTTATATAGAATACTATATATATATAAATATATATATATATAT	18208	
Qy	3030	TAATGATGTTTAAATTTAAATATATAATGATGCCGGATATTTTATTTTAAAGTACTCTGG	3089	
Db	18207	TATATATTAATTTAAATAATTTATATAATGATTTATATAAATAATATATAAATAAT-A	18149	
Qy	3090	TTGATTTATTTTATTCAGAGCAAAAGTAGTAGTATTTTATTAATTTTATGATTTAAAT	3149	
Db	18148	TAATTTTATTTATATAAATTTATATAAATTAATTAATATAAATAATATAAATAAATAAT	18089	
Qy	3150	TTATTTTCATATAAAATTTTGGAAACTAACTAAATATAAAAAATGAAATTTTATGTGTT	3209	
Db	18088	ATATATATCATATATGTTATATAAAAAATATAAATATATATAAATAATGTAATGT	18029	
Qy	3210	TTATTTAGGCTATATATGTTTTGTTTCCAGTAGTCACAAGTAGTGTGTTGGAATAAAT	3269	
Db	18028	ATAATATATATAATATATATATATATATATATATATAAAAAATATAAATATATATATAA	17969	
Qy	3270	TTTGGAAAGATTATTCAGAT--TTTACTGCTCCCATAAATTCGATTTTGCATAATGTA	3328	
Db	17968	TAGTATATAAATTTCAATATAATATAAATAATATCATATATAAATAAATAATATAA	17909	
Qy	3329	TTATAATTTGTATTTCAATTTATAAATATGATTTATAAAAAATTTAAAAATATGTTCTTTT	3388	
Db	17908	ATATATATCATATAATATAATATATATATATATATATATATATATATATATATATAT	17849	
Qy	3389	TAGTTTTTTAGTTTTATTTAGGTATATCTGGATTTGTATATTTATTCAAAATGGGAAAGATAT	3448	
Db	17848	ATATTATATATATATAATAATAATAATATATATATATATATATATATATATATATATAT	17789	
Qy	3449	TGATTTTTTAGACAGACACCTTATAGCATAGACTATCTTATA-ACAGCGCGTCAAAACAA	3507	
Db	17788	TATATATATATAATACATATAATATAATAAATAAATAAATAATATATATATATATATAC	17729	
Qy	3508	GGTTGGTTGGCTTTATAGCAATCTCCCTAGTTAAATACCACTACAATATATAGTTTCCAAATTC	3567	
Db	17728	ATTATATATATATATATACATTAATATATATATATATATATATATATATATATATATAT	17669	
Qy	3568	CGTTAACTCTTGCACTTATAAAAAATAAATGCAACAAATTTTTTTCTTGTGCTGTGCTT	3627	
Db	17668	ATTATATATATATATATATATATATATTTTATATATATATATATATATATATATATATAT	17609	
Qy	3628	TTATACCGACTCTATTTAAGTGGATCGAGAAATGGTAGTTTATTCGTAGCAATTAATTA	3687	
Db	17608	TA	17553	
Qy	3688	TATGCTTTGTTATCGAGATATATAGGTGGAAAAATTTTCTTTGGATAAAAAGCTAAATAGTAA	3747	

## RESULT 12

```

US-09-489-039A-1006
Sequence 1006, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND
TITLE OF INVENTION: PNEUMONIAE FOR DI
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,7
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 1006
LENGTH: 1521
TYPE: DNA
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1006

```

Query Match	1.9%	Score 132.2;	DB 4;	Length 1521;
Best Local Similarity	54.1%;	Pred. No. 8.5e-15;		
Matches 340;	Conservative 0;	Mismatches 278;	Indels 11;	Gaps 3;
Qy	746	AAATTTTATAAGCCCTAGTCATATCATGATGAAACGAC	TTTTCGGATATATCTCGGAGCGGT	805
Db	897	AAATTTAGCTTAATGTCATCGAATTCCTTAAGCGTACAT	TTTTCATATTTCTGGGCTCT	956
Qy	806	AGTCGGGTTAATTTATTTGGGTATAGTTTCTATTTTGT	TTTCCCAATTTATTCGTAGAGA	865
Db	957	GGCAATAAATTCCTCTTTTCGGCATTATTAATTTAT	TTTAGTAAAGGTAAGAAAGGA	1016
Qy	866	TGCTGGACCGGCTATTTTGTCTCAGAAAACGAGTTGG	ACAGAAATGGACGATATTTACATT	925
Db	1017	CGCGGTACAGCTATATACGGGATCGAGCTGTAGTA	AGGGTGGTAAACCATTTAAATC	1076
Qy	926	CTACAAGTTTCGATCGATGTATGTTGATGCTAGGAG	CGCAAAAAGACATTTGTC-----	980
Db	1077	CCTCAAAATTCGGTTCAATGGTTGTAATTTCTAAGA	GAAGTATTATAGAAGAGCTGTGGCANA	1136
Qy	981	--AGCCAAAACCAAGATGCAAGGGTGGGTATGTTTT	AAAAATGGCAAAAACGATCCTAGAAT	1038

Db 1137 TGATCCAGAGTAAAGCAGAAATGGGATGCAACTTTTCAAACTGAAAAACGATCCTCGCAT 1196  
Qy 1039 TACTCCCAATTCGACATTTTCATACGCAAAAAAAGTTTAGACGAGTTACCAACAGTTTATATA 1098  
Db 1197 TACTAGTATTGGTCATTTTACGTA-GGACAAGTTGGATGAACCTTCTCAATTGTTCA 1255  
Qy 1099 ATGTTTTAAATGGCGATAGAGTCTAGTTGGTACACGTCACCTACAGTTGATGAATTTG 1158  
Db 1256 ATGGCTTTAAAGGGGAGATGAGTTTAGTAGGCCACCGCCGATCATTTACACGGGAGTTAG 1315  
Qy 1159 AABAATATCTCCCTGGTCAAAGAGACGATTTAGTTTAAACCGAGGATTTACAGTCTCT 1218  
Db 1316 AGCGATATAACGATGAGTCCGACTACTATCTATTAAAGTAAACCTGGGATGACTGGGCTAT 1375  
Qy 1219 GCGAGGTAGTGGTCGTAGTAAATACACAGACTTCGACGAGTGTAGTTCCGTTGGACTTAG 1278  
Db 1376 GCGAGGTAGTGGTCGTAGTAAATACACAGACTTCGACGAGTGTAGTTCCGTTGGACTTAG 1432  
Qy 1279 CATACATTGATTAATTTGGACTATCTGGTCAGATATTTAAATTTTAAAGACAGTGAAAG 1338  
Db 1433 GGTATGTAATAAACTGGTCTATGTGGAATGATATCGCAATCTTGTTCAAAACCATAAATG 1492  
Qy 1339 TTGTTATTTGTCAGAGAGGAGTAAAGTAA 1367  
Db 1493 TTGTTATTTAAAGAGATGGCGCTTATTGA 1521

## RESULT 13

US-09-949-016-14164  
; Sequence 14164, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASES, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: C1001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14164  
; LENGTH: 18773  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-14164

Query Match 1.8%; Score 129.2; DB 4; Length 18773;  
Best Local Similarity 45.3%; Pred. No. 4.9e-14;  
Matches 595; Conservative 0; Mismatches 708; Indels 11; Gaps 3;  
Qy 2090 TTTCTAGACAAAAAGATGGTGAACATGTAATGATCATCAAGTAGAGTTGTGAAGA 2149  
Db 17099 TTTCTTTATTTCTATATATATAATATTTTATATATAATATATATATAATATTTTA 17158  
Qy 2150 AGAATTTTACAAGATAATAATTTTATTTATAGAAAAATAGATGATTTTGTGAAAAA 2209  
Db 17159 TATATAATTTTATATATAAATATATATATAATATATATAATATTTTCTATATAA 17218  
Qy 2210 ATTATGAGTTTCTTAAGCAAACTAATTTTACATCAATAATAATTTTGTGGAAGA 2269  
Db 17219 ATGGTATATAATTTATATAATTTATATAAATAATATATAGATATCTAATAATGTATA 17278  
Qy 2270 TTAACAATAAGTTGAAAAATTTTAAAGAGATCAAGAAATGATTAATAAAAAAGATGC 2329  
Db 17279 ATATATAACATATAAAAAATATTTTAAATATATAATATTTTATATATAATATTTTAT 17338

Qy 2330 ATATTGATTAATGGCTTATCAATAATTTTCTCAGATTTTACTGGAGAGGATACAGATAT 2389  
Db 17339 ATATAATATAATATATATTTTATATAAATTTATTAATATAATATAATATAATATAAT 17398  
Qy 2390 TATCATCTTCTCAGGAGATGCACACCAATTTAGTCTCTTCAGAAATCCTCTATAATAT 2449  
Db 17399 ATTTTATACATAATTTATTTATATAATTTATATAATATAATATAATATAATATAAT 17455  
Qy 2450 TTTAAATATTTCTCAGGATTTTATATGTTGAAATTTTACAAAAGATGAGCAAAAAATATAAGAA 2509  
Db 17456 TTATCAATTTATATAATAATTAATATAATATAATATAATATAATATAATATAATATA 17515  
Qy 2510 AATAGGATATATGAACAGTTAAATGTACAGATTTTCTTAATATATACAGAAAAAAT 2569  
Db 17516 TATATTTATTTATATAATATAATATAATATAATATAATATAATATAATATAATATA 17575  
Qy 2570 ATTGATTAATGACTGTTTGAATTTTATTAAGAAATGATCGAGCTTTTGAATACTATT 2629  
Db 17576 GTATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAAT 17635  
Qy 2630 CAAAAGATTTGTTTATTTAGTAGAATAAAAAACATGGTCTAAAGAAATGAATTTGGTTCTA 2689  
Db 17636 TAAAAATATAATATAATATAATATAATATAATATAATATAATATAATATAATATA 17693  
Qy 2690 ATTGGGTTTCGCTCCACATGATTTTGTGGCAATTTCTTTATCAAAATGAAACCGAAACAG 2749  
Db 17694 ATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAATATA 17753  
Qy 2750 CTTATTTTAAAGTAACTTAAATGTCAGAGTAACTTTTATACAGCAATTTATAGAA 2809  
Db 17754 ATATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAAT 17813  
Qy 2810 AATATGAATTTTCAAAATAGATTTCTAAATATGAAAAATTTAAAGATATAATAAGTGA 2869  
Db 17814 ATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAATATA 17873  
Qy 2870 AATCAACATCTTCTCTATTTGCTTTACAGATGATTTCTTTATGAGAAATGCTAAATGCA 2929  
Db 17874 TATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAATATA 17933  
Qy 2930 GAAATTTTGGTTTATTTGCTAGAAAGTTTAAATAAGAAAAATATAATATAATATAAG 2989  
Db 17934 TATATTTTATATAATATAATATAATATAATATAATATAATATAATATAATATAAT 17993  
Qy 2990 AAATTTTAAATAAATAAATAGTTGATTTTCTGAGAGTAAATGATGTTTAA----- 3043  
Db 17994 ATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAATATA 18053  
Qy 3044 ATTTTAAATAATAGACCCGGAATTTTATTTTAAAGTACTTCTGGTTGATTTATTTTAT 3103  
Db 18054 ATTATAATTTTATATAACAATATGATATAATAATAATATAATATAATATAATATA 18113  
Qy 3104 TCCGAGCAAAAGTATGATTTTATTTTAAATTTTATGAATTTTATTTTATTTTATATA 3163  
Db 18114 TATAATTTTATATAAATTTATATAAATAAATAAATTTATATAATATAATATAATATA 18173  
Qy 3164 ATTTTCAAACTAAAGCTAATAATTTAAATAAATTTTATTTTATTTTATTTTATTT 3223  
Db 18174 AATCATTTATATAATTTATATAATTTATATAATATAATATAATATAATATAATATA 18233  
Qy 3224 ATTATGTTTGTTCAGTAGTCACAGTATGTTTGTGAAATAAATTTTGAAGATTTAT 3283  
Db 18234 ATAGTATTTCTATATAAATAATAATAAATAATTTTATATAAGAAATTTATATAATATCA 18293  
Qy 3284 TGCAGATTTTACGCTCCCAATAATTTGGATTTATTTGCAATTAATGTTATTAATTTGTTATTC 3343  
Db 18294 TATATAATTTTATAGAAATTTATATAATAATAATATAATATAATTTTATATAAGAAATTTTA 18353  
Qy 3344 ATTTTAAATATTCGATTTATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 3397  
Db 18354 TATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAATATA 18407



```

RESULT 14
US-09-583-110-893
; Sequence 893, Application US/09583110
; Patent NO. 669703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 893
; LENGTH: 987
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; US-09-583-110-893

```

Query Match.	1.8%;	Score 127.4;	DB 4;	Length 987;
Best local Similarity	51.2%;	Pred. No. 5.7e-14;		
Matches 327;	Conservative 0;	Mismatches 306;	Indels 6;	Gaps 1;
QY	6173	AAATTCCTATAATGTGTACCTATATATAATGTAGAAATAATTTATCTAAATCTATAGATAG	6232	
Db	15	AAATAGTGTGATGTGTGCAGTCTATAATGTGGCGCAGTACCTAGAAANATCGATAGTTC	74	
QY	6233	CAATGTAAATCAGACCTCAACAACATATAGAGATTCCTCTGGTGAATGACGCTAGTACGGA	6292	
Db	75	CAATTCAGAAGCAGACCTATCAAAATCTGGAATTTATTCCTGTGATGATGGTGCACAGA	134	
QY	6293	TAATTCGGAAGAAATTTGTTTAGCATATCGAAGAAGATAGTCGATTCGTTATTTTAA	6352	
Db	135	TGAAGTGGTCGTGTGTGATTCAAATCGCTGAAACAAGATGACAGGGGTGCATGTCCTCA	194	
QY	6353	AAAAGAGAAGCGCGGGCTATCAGATGCCGTAAATTTATGGCAATAGTCGCGCCAAGGGTGA	6412	
Db	195	TAAAGAAACGAAGGATTTGTCGAAGCACGAANATGATGGATGAAGCAGGCTCACGGGA	254	
QY	6413	CTACTTAGCTTTTATAGACTCAGATGATTTTATTCATTCGAGTTCATCCAACTTTTACA	6472	
Db	255	TTATCTCAATTTTATTGACTCAGATGATTTATCCATCCAGAAATGATTCAGAGCTTATA	314	
QY	6473	CGAAGCAATTGAGAGAGAGAAATGCCCTGTGTGGCAATGCTGTGTATCATAGGGTAGATGC	6532	
Db	315	TGAGCAATTAGTTCAGAAGATGC-----GGATGTTTCGAGCTGTGGTGCATGATGT	368	
QY	6533	TTCGGGCATTTCTTAAACAGCAGAGCGCTTCCTCAAAATCAGGCTGTTCTGAGCGGCAG	6592	
Db	369	CTATGCTAATGATGAAGACCCACAGTCAGCCCAATCAGGATGACTATTTGTGTGTATTC	428	
QY	6593	GAATGTTGTGAAAAGCTGCTAGAGGGGATGGTTCATCGCTTTGTGGTGGCTGTAATAA	6652	
Db	429	TCAACAATTTCTAAGGAATACCTCATAGGTGAATAAATACCTGGGACGATTTGCATATA	488	
QY	6653	ACTCTATAAAAAGAACTATTGTGAAGATTTTCGATTTTGAAGAAGGTTAAGATTCATGAAGA	6712	
Db	489	GCTAATCAAGAGACAGATTCGAACGTGACCTATCCCTTCTTAAGGGTGTGATTTACGAAGA	548	
QY	6713	TGAATCTTCACCTATCCGTGTGCTCATGAGTTAGRAAAGATTCGAATAGTTAAGGATG	6772	
Db	549	TGCCTATTACATTTTGAATTTAATCAAGTTGGCCAGAAGTATGTGGTTAATCTAAACC	608	
QY	6773	CTTGTACTATTATGTTCACCGAGAAATAGTATCACAAAC	6811	
Db	609	CTATTTATTAATAATTTTCATAGAGGGGATGATTTACGAC	647	

```

RESULT 15
US-09-107-433-2107
; Sequence 2107, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
;
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
;
; INFORMATION FOR SEQ ID NO: 2107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 996 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
;
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
;
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...996
; SEQUENCE DESCRIPTION: SEQ ID NO: 2107:
US-09-107-433-2107

```

Query Match	1.8%	Score	127.4;	DB 4;	Length	996;
Best Local Similarity	51.2%;	Pred. No.	5.8e-14;			
Matches	327;	Conservative	0;	Mismatches	306;	Indels
					6;	Gaps
					1;	
Qy	6173	AAATTTCTATAATGTACCTATATATAATGTAGAAAAATATTTATCTATAAATGTATAGATAG	6232			
Db	24	AAATAGTGTGATTTGCGCAGTCTATATATGTGGCGCAGTACCTTGAATAATTCGATAGCTTC	83			
Qy	6233	CAATTTGAATFCAGACCTCACAAACATATAGAGATTCCTCTGGTGAATGACGGTAGTAGCGGA	6292			
Db	84	CAATTCAGACGACACCTATCAAAATCTGSAATTAATTCCTGTTGATGATGGTGCACACAGA	143			
Qy	6293	TAATTCGGAGAAATTTGTTTACCATATCGAGAGAAGATAGTCGCATTCGTTATTTTAA	6352			
Db	144	TGAAGTGTGCGCTTGTGTGATTCATTCGCTGAACAAGATGACAGGGTGTGAGTGCTTCA	203			
Qy	6353	AAAAGAGAACGGCGGGCTATCAGATCCCGTAATATTGGCAATAAGTCGCGCCCAAGGGTGA	6412			
Db	204	TAAAAAGAACCAAGGATTTGCGCAAGCACGAAATGATGGATCAAGCAGGCTCACGGGGA	263			





Db 4107 TTATCTATCGTATATAAAATTAATTTATTTATTAATAATTAACAATAAGTGAAGAAATCTTAG 4048  
Qy 6070 AAAATTCCTTCTCAAAAATTTTGTATPAAGAAATTTGTTTGAACAAGCTTTTAAAAAAA 6129  
Db 4047 ATATTTTATATATATATATCTTTTATATAACAATAAATTCAGATTCCTAAATGAGGAA 3988  
Qy 6130 TATTATGTTTATATAGGAAGATCAAGGATATCTATAGTAAATTTCTATATTTGTAC 6189  
Db 3987 ATATATATTTGTATTAATGGAAGCTTAATGGTCCATATTAATAAATTAATAACAACCTT 3928  
Qy 6190 CTA-TATATAATGTAGAAAAATATTTATCTAAATGTATAGATAGCAT 6235  
Db 3927 ATATTAACAATTTTATATAATAATATAAACAACAACAATAAAT 3881

RESULT 17  
US-09-949-016-17067/c  
; Sequence 17067, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17067  
; LENGTH: 95255  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(95255)  
; OTHER INFORMATION: n = A, T, C or G  
US-09-949-016-17067

Query Match 1.8%; Score 125.8; DB 4; Length 95255;  
Best Local Similarity 45.6%; Pred. No. 2.8e-13;  
Matches 903; Conservative 0; Mismatches 1047; Indels 29; Gaps 12;  
Qy 2229 AAACCTTACATCAAAATTAATTTTGTGAAGATTAACAAATAGTTGAAA 2288  
Db 43924 AAATGAGTTTTCACCTAAAAATCTGAATTTCCGGGGCTTCTTTGAAAAAAATTAGTA 43865  
Qy 2289 AATTATGAGGATCAAGAAATGAATATAAAGATGATATTTGATTAATGCTTAT 2348  
Db 43864 TCTGTGATACATCTATTTCTGCATACATCTACTGAGCTGGATCTGTGGCTGACCTCTTT 43805  
Qy 2349 CATAATTTTCTCAGATTTTACTGAGAGGATACAGATATTTATCATCTTCTCTCAGGAG 2408  
Db 43804 TAGAAGGCGATGCCATTTTGTAGTCCGACCTGCTCCCTGTCATTCGGAG 43745  
Qy 2409 AATGCACACCATTAAGTTCTCTTTCAGAAATACCTGTATATTTTAAATTTCTCAGGAT 2468  
Db 43744 ACTGATCCC---AGGCCCTCTCATCTTATAATTTCTTCTTATAATTTATTATA 43689  
Qy 2469 TATATGTTGAATTTACAAAGATGAGCAAAATATAAAGAAATAGGATATATGACGAG 2528  
Db 43688 TAT 43629  
Qy 2529 TTAATGTTTACAGATTTTCTTCTTAATATATATATATATATATATATATATATATAT 2588  
Db 43628 TCATATATAATTTCTATACAT---TTGTATATATATATATATATATATATATATATAT 43570  
Qy 2589 GAAATTTTAAAGATGTATCGAGCTTTTGAATCTATTTTAAAGATTTGTTTATTG 2648

Db 43569 TATAATTTATAAATAATATACAAATTTCTATATTTATATATCTTTGTATACAAATATACAAAT 43510  
Qy 2649 ATAGAATAAAAAAACAATGGTCTTAAGAATAAGATTTGGTTCTTAATTTGGGTTTCGCTTCCACA 2708  
Db 43509 CTATATCATAGTTTGTATATAATATAAATCTCTGTATATCATATCTTTGTATATAATATA 43450  
Qy 2709 TGATTTTGTGGCAATCTTTTATCAATGAAGAAACGAAACAGCTTATTTATTTAGTAATC 2768  
Db 43449 CAATTTCTATGTATATCATATCTTTGTATATAAATACAAATTTCTGTATATCATATTTTGT 43390  
Qy 2769 TAAATGTCAGATGAACCTTATTTATACAGACAAATTTAGAAAAATATGAATTTTCAAAATAG 2828  
Db 43389 ATATAATATACAATCTGTATATTTTGTATATAATATACAAATTTCTGTATATATCATAT 43330  
Qy 2829 ATATCTAAATATGGAAT---TTAAGATATATAAAGTGAAGAAATCAACATCTTCTC 2884  
Db 43329 TTTGTACATAATATACAAATCTGTATATCATATTTTGTATATAATATACAAATCTCTGT 43270  
Qy 2885 CTATTTGCTTT---ACAGATGATTTCTATTTGATGAATTTGCTAAATGCAAGAAATTTAGTT 2941  
Db 43269 ATATTTTGTATATAATATACAAATCTGTATATATTTTGTATATAATATACAAATCTCTGT 43210  
Qy 2942 TTTTATTTGCTAGAACTTAAATAAGAAATAAATCTA---AATTTAAAGAAATTTATTA 2998  
Db 43209 ATATTTTGTATATAATATACAAATCTGTATATTTTGTATATAATATATAATTTCTGTAT 43150  
Qy 2999 CTAAAAAATAAATAGTTGATTTTGTGAGAGTAATGATGTTTAAATTTTAAATTTTAAATATGA 3058  
Db 43149 ATTTTGTATACATATATAATCTGTATATATTTTGTATATAATATATAATTTGTTA 43090  
Qy 3059 CCGGAATATTTTATTTTAAAGTACTTCTGCTGATTTATTTTATTCAGAGCAAAAGTA 3118  
Db 43089 TATTTTGTATATAATATAAATTAATATATATATAATTAATTAATTTATTAATATA 43030  
Qy 3119 TGATTTTATTAATTTTATGAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 3178  
Db 43029 TAATATATAAATTTATTAATATAATATAATATAATATAATATAATATAATATAATATAAT 42970  
Qy 3179 GCTAATATAAATAAGAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 3238  
Db 42969 TAAATATATAATTTCTATATTTATTTTGTATATAATATAATATAATATAATATAATATA 42911  
Qy 3239 AGTAGTCACAGTATGTTTGTGAATAAATTTTGAAGATTTATTTGAGATTTTACTGTC 3298  
Db 42910 ATAATCTATATTTATGTTATATTTTGTATATAATATAATATAATATAATATAATATAAT 42851  
Qy 3299 TCCATTAATTTGGATTTATGCAATATGATTAATTTGATTTGATTTTATTAATATAATTTGA 3358  
Db 42850 TATATTATATTTTGTATATAATATAATATAATATAATATAATATAATATAATATAATATA 42791  
Qy 3359 TTATAAAAAATTAATAATAGTATCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 3418  
Db 42790 ATATAATATAAATCTGTATATTTTGTATATAATATAAATCTGTATTTATTTATTTATTT 42731  
Qy 3419 ATTGTATATTTTCAAAATGGGAAGATTTTGTATTTTATTTTATTTTATTTTATTTTATTTTAT 3478  
Db 42730 TGTATATAATACAAATCTGTATATTTTGTATATAATATAATATAATATAATATAATATAAT 42671  
Qy 3479 AGACTATCTTATACAGGCGTCAAAACAAGGTTGGTGGCTTTATTAAGAACTAT-CCTAGCT 3537  
Db 42670 TGTATATACCATTTGTATATTTTGTATATAATATAAATTTGTATATTTTGTATATAATA 42611  
Qy 3538 TAAATACCATACAAATTTATAGTTTCAATTCGGTTAACTTTGCACTTATTTAAGTCGATCGAGAA 3597  
Db 42610 TATAATTTCTGTATTTATATTTTGTATATAAATATAAATTTCTGTATATTTATTTTGT 42551  
Qy 3598 TGCAACAATTTTCTTGTCTTCTGTTTATACCGATCTATTTAAGTCGATCGAGAA 3657  
Db 42550 TATATATAATATAATTTCTGTATATTTATTTT---TGTATATAATTTCTGTATATTTAT 42494  
Qy 3658 TTGTAGTTTATCGCTAGCAATATTAATTTATCTTTTGTATATGAGATATATAGTGGAA 3717







Qy	6658	ATAAAAGAACTATTATTGAAGATTTTCGATTTGAAAGGGTAAGATTTCATGAAGATGAAT	6717
Db	8723	ACCCGGTTCTTTCTTTTGAACAATTTCTTTCCAGAGGAAACATTTACGAAGATATGG	8782
Qy	6718	ACTTCACCTTATCGGTTTGCTCTATGAGTTAGAAAAAGTTGCAATAGTTAAAGAGTGCTTGT	6777
Db	8783	GAACAACTTACAAATTAATAAAATTGCAAGTGAAGTGGTCTTCTTGGATCGGTATGATT	8842
Qy	6778	ACTATTATGTTGACCGAGAAAATAGTATCACAAC	6812
Db	8843	ATGCCTACGTACAGCGCACTTAATAGTATCATGAT	8877

## RESULT 21

```

US-09-495-406-1/c
; Sequence 1, Application US/09495406
; Patent No. 6503744
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000110US
; CURRENT APPLICATION NUMBER: US/09/495,406
; CURRENT FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 11474
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: 11.5 kb PCR product from C. jejuni. OH4384
; OTHER INFORMATION: including LOS biosynthesis locus
US-09-495-406-1

```

Query Match 1.7%; Score 121.4; DB 4; Length 11474;

	Best Local Similarity	54.8%; Pred No. 1.1e-12;	Mismatches 0; Gaps 4;
	Matches 324; Conservative		
QY	4998 TAAATGAGTTCTTCGCCAATATTAGAGTCTTATCGAAAAGAATAACGTAGATATCCATTAT	S057	
Db	6132 TAAATACATCAAAATCATTTTGTTAGTCTTGAATTAATCAATTTCTTTTAAACTTGGTCCAT	6073	
QY	5058 TAAAGCGAAAGATATTTATCAAGAAGCAATTTAGTTTACGTTGTATTTGATGAATAATTTTC	5117	
Db	6072 TTCCAGCAATAAACAATTTTTTTCATGATTTATTTTCCATCAAGATTATTTTTTAAACGATTT	6013	
QY	5118 GCCTTAACTATATGTTAATGTTTATATAAGAAATTTCAAACGAGTAGAGGTAAAAATCGAT	5177	
Db	6012 TATCAAANTAGATTTTATGTTATA - ATTAATTTATATAAANAATGAAGAAATAAATATGTTT	5954	
QY	5178 AAAATTAGTCTTATTTGTTCCAGTTTATAATGTTAGATATAATTTTAAGTAGTCTGTATAGAA	5237	
Db	5953 AAAATTCCAATCATCTTACCACCTTATATCTCGAACATATATAGCAGGGCANATAGAA	5894	
QY	5238 AGCATTTATTAATCAAAATTTATAAAAATATAGAAATATTTTGTATAGATGATGCTCTGTA	5297	
Db	5893 AGCTGTATCAATCAGACTTTTAAAGATATAGAAATAATTTGATGTTGATGTTGGAAT	5834	
QY	5298 GATGATCTCTCAAAATATCGAAGGAATATGCAGAAAAAGATATAAGAGCTATAAAAATTTTT	5357	
Db	5833 GATTAATAGTATAAATATAGCCAAGAAATCTCTTAAAAAAGACAAAAAGAAATAAAAAATAATC	5774	
QY	5358 TTCACAT - --AATCATAGTGGAGTATCAAAATGCTAGAAATCATCGAATAAAGCGGAGTACA	5414	
Db	5773 CACAATGMAAAAAAATCTTAGTCTTTTAAAGCAAGATATGAAGGTGTGAANGTAGCAATC	5714	
QY	5415 GCTGAATATATTTATGTTTGTGACTCTGATGATGTTGTTGATAGTAGATTTAGTAGAAAAA	5474	

## RESULT 22

```

US-09-816-028A-1/C
; Sequence 1, Application US/09816028A
; Patent No. 6699705
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/09/816,028A
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 11474
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384
; OTHER INFORMATION: including LOS biosynthesis locus
US-09-816-028A-1

```

Query Match 1.7%; Score 121.4; DB 4; Length 11474;

	Best Local Similarity	54.8%;	Pred. No. 1.1e-12;	Indels	6;	Gaps	4;
	Matches	324;	Conservative	0;	Mismatches	261;	
QY	4998	TAATGAGTCTCTGCCAATATATTAGAGCTCTTATCGAAAGAGAAATACGTAAGATATCCATTAT					5057
Db	6132	TAAATACATCAAAATCATTTGGTAGTCTTGAATAATCAATTTCTTTTAACTTGGTCCAT					6073
QY	5058	TAAAGCGAAAGANATTTTATCAAGAAAGCATTTAGTTTACGTTGTATTTGATGAAATTTTC					5117
Db	6072	TTCCAGCAATAATAACTTTTTTTCATGATTATTTTCCATCAAGATTTATTTTAAACGATTT					6013
QY	5118	GCCTAAACTATATGTAACTGTATATAAGAAATTTCAAAGCAGCTAGAGGTAAAGAAATGGAT					5177
Db	6012	TATCAAAATAGAGTTTATGTGTATA-ATTAAATTTATATAAATAAAGCAATAATATGTGTTT					5954
QY	5178	AAAAATTGAGTGTTATCTGCTCCAGTTTATAATGTAGATAAATATTTTAAGTAGTTTCTGTAGAA					5237
Db	5953	AAATTTTCAATCACTTACCAACTTATTAATGTGGAAACAATATATAGCAAGGGCAATAGAA					5894
QY	5238	AGCATTTATTAATCAAAATTTATAAAAATATAGAAATATTATTGATAGATATGCTCTGTA					5297
Db	5893	AGCTGTATCAATCAGACTTTTTAAAGATATAGAAATAATTTGTAGTTGATGTTGGGAAAT					5834
QY	5298	GATGATCTTGCTTAAATATGCAAGGAATATGCGAAAAAGATAAAGAGGTAAAGATTTTTT					5357
Db	5833	GATAATAGTATAAATATATGCCAAAGAAATCTCTAAAAAAAGACAAAAAGATAAAAAATAATC					5774
QY	5358	TTCACT---AATCATAGTCGAGTATCAATGCTAGAAATCATCGAATTAAGCCGAGTACA					5414
Db	5773	CACAATGAAAAAAAACCTTAGGTCCTTTTAAAGACAAGATATGAAGGTGTGAAGGTAGCAAAC					5714
QY	5415	GCTGAATATATTTATGTTTGTGACTCTGATGATGTTGTTGCTAGTAGTAGATTAGTAGAAAAA					5474



Db 5713 TCCTCTATATAATGTTTGTAGCTCTGATGATTAATTGGA-CTAAATGCTGTGTGAGA 5655  
Qy 5475 TTATATTTTAAATATAAATAAGTAGAAGTATTAATCTCGT-TGTTTGTACGCTACTTT 5533  
Db 5654 GTGTATAAAAATTTTAGATGAACAGGATGAAGTTGATTTAGTGTGTTTTCATGCTATTGT 5595  
Qy 5534 TTCAGAAATATAAATATTTTGAAGTGAATAATCCAAATATTCATTTTGA 5584  
Db 5594 TGAAGTAATGTTATTTTCATATAAAAAAGTTTGACTTTTAATTCGTGTTTTTA 5544

## RESULT 23

US-10-303-162-1/c  
; Sequence 1, Application US/10303162  
; Patent No. 6723545  
; GENERAL INFORMATION:  
; APPLICANT: Gilbert, Michel  
; APPLICANT: Wakarchuk, Warren W.  
; APPLICANT: National Research Council of Canada  
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of  
; FILE OF INVENTION: Gangliosides and Ganglioside Mimics  
; FILE REFERENCE: 019633-000111US  
; CURRENT APPLICATION NUMBER: US/10/303,162  
; CURRENT FILING DATE: 2002-11-21  
; PRIOR APPLICATION NUMBER: US/09/816,028  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: US 60/118,213  
; PRIOR FILING DATE: 1999-02-01  
; PRIOR APPLICATION NUMBER: US 09/495,406  
; PRIOR FILING DATE: 2000-01-31  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 11474  
; TYPE: DNA  
; ORGANISM: Campylobacter jejuni  
; FEATURE:  
; OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384  
; OTHER INFORMATION: including LOS biosynthesis locus  
US-10-303-162-1

Query Match 1.7%; Score 121.4; DB 4; Length 11474;  
Best Local Similarity 54.8%; Pred. No. 1.1e-12;  
Matches 324; Conservative 0; Mismatches 261; Indels 6; Gaps 4;  
Qy 4998 TAATGAGTTCTGCGCAATATTAGAGTCTTATCGAAAAGAAATAGTAGATATCCATTTAT 5057  
Db 6132 TAAATACATCAAAATCAATTTGGTAGTCTTGAATAATCAATTTCTTTTAACTGGTCCAT 6073  
Qy 5058 TAAAGCGAAAAGATATTTATCAAGAAAGCATTTAGTTACGTTGTATTTGATGAAATTTTC 5117  
Db 6072 TTCCAGCAATAATAACTTTTTCATGATTAATTTTCCATCAAGATTTATTTTAAACGATTT 6013  
Qy 5118 GCCTAAACTATATGTAATGTTATATAAGAAATTTCAAAGCAGTAGAGGTAAGAAATCGAT 5177  
Db 6012 TATCAAAATAGATTTTATGTTATA-ATTAAATATAATAAAATAGGAATAAATATGTTT 5954  
Qy 5178 AAAATTAGTGTATTGTTCCAGTTTATTAATGTAGATAAATATTTTAACTAGTTGTATAGAA 5237  
Db 5953 AAAATTTCAATCATCTTACCAACTTATAATGTGGACAAATATATAGCAAGGCAATAGAA 5894  
Qy 5238 AGCATTTAATCAAAATATAAATAATAGAAATATTTTGAATGATGATGCTCTGTA 5297  
Db 5893 AGCTGTATCAATCAGACTTTTAAAGATATAGAAATAATTTGATGATGTTGGAAT 5834  
Qy 5298 GATGATCTGCTAAATATGCAAGGAATATCGAAGAAAGATAAAGAGTAAGAAATTTT 5357  
Db 5833 GATTAATAGTATAATATAGCCAAAGATATCTTAAAGAAAGCAAAAGAAATAAATAATC 5774  
Qy 5358 TTCACT---AATCATAGTGGAGTATCAAAATGCTAGAAATCATGGAATAAAGCGGAGTACA 5414  
Db 5773 CACAATGAAAATAAATCTTAGGTTCTTTTAAAGCAAGATATGAAGGTGTAAGAGTACCAAC 5714

Qy 5415 GCTGAATATATATGTTTGTGACTCTGATGATGTTGTTGATAGTAGATAGTAGAAAAA 5474  
Db 5713 TCCTCTATATAATGTTTGTAGATCTGATGATTAATTGGA-CTAAATGCTGTGTGAGA 5655  
Qy 5475 TTATATTTTAAATATAAATAAGTAGAAGTATTAATCTCGT-TGTTTGTACGCTACTTT 5533  
Db 5654 GTGTATAAAAATTTTAGATGAACAGGATGAAGTTGATTTAGTGTGTTTTCATGCTATTGT 5595  
Qy 5534 TTCAGAAATATAAATATTTTGAAGTGAATAATCCAAATATTCATTTTGA 5584  
Db 5594 TGAAGTAATGTTATTTTCATATAAAAAAGTTTGACTTTTAATTCGTGTTTTTA 5544

## RESULT 24

US-10-303-134-1/c  
; Sequence 1, Application US/10303134  
; Patent No. 6825019  
; GENERAL INFORMATION:  
; APPLICANT: Gilbert, Michel  
; APPLICANT: Wakarchuk, Warren W.  
; APPLICANT: National Research Council of Canada  
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of  
; FILE OF INVENTION: Gangliosides and Ganglioside Mimics  
; FILE REFERENCE: 019633-000111US  
; CURRENT APPLICATION NUMBER: US/10/303,134  
; CURRENT FILING DATE: 2002-11-21  
; PRIOR APPLICATION NUMBER: US/09/816,028  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: US 60/118,213  
; PRIOR FILING DATE: 1999-02-01  
; PRIOR APPLICATION NUMBER: US 09/495,406  
; PRIOR FILING DATE: 2000-01-31  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 11474  
; TYPE: DNA  
; ORGANISM: Campylobacter jejuni  
; FEATURE:  
; OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384  
; OTHER INFORMATION: including LOS biosynthesis locus  
US-10-303-134-1

Query Match 1.7%; Score 121.4; DB 4; Length 11474;  
Best Local Similarity 54.8%; Pred. No. 1.1e-12;  
Matches 324; Conservative 0; Mismatches 261; Indels 6; Gaps 4;  
Qy 4998 TAATGAGTTCTGCGCAATATTAGAGTCTTATCGAAAAGAAATAGTAGATATCCATTTAT 5057  
Db 6132 TAAATACATCAAAATCAATTTGGTAGTCTTGAATAATCAATTTCTTTTAACTGGTCCAT 6073  
Qy 5058 TAAAGCGAAAAGATATTTATCAAGAAAGCATTTAGTTACGTTGTATTTGATGAAATTTTC 5117  
Db 6072 TTCCAGCAATAATAACTTTTTCATGATTAATTTTCCATCAAGATTTATTTTAAACGATTT 6013  
Qy 5118 GCCTAAACTATATGTAATGTTATATAAGAAATTTCAAAGCAGTAGAGGTAAGAAATCGAT 5177  
Db 6012 TATCAAAATAGATTTTATGTTATA-ATTAAATATAATAAAATAGGAATAAATATGTTT 5954  
Qy 5178 AAAATTAGTGTATTGTTCCAGTTTATTAATGTAGATAAATATTTTAACTAGTTGTATAGAA 5237  
Db 5953 AAAATTTCAATCATCTTACCAACTTATAATGTGGACAAATATATAGCAAGGCAATAGAA 5894  
Qy 5238 AGCATTTAATCAAAATATAAATAATAGAAATATTTTGAATGATGATGCTCTGTA 5297  
Db 5893 AGCTGTATCAATCAGACTTTTAAAGATATAGAAATAATTTGATGATGTTGGAAT 5834  
Qy 5298 GATGATCTGCTAAATATGCAAGGAATATCGAAGAAAGATAAAGAGTAAGAAATTTT 5357  
Db 5833 GATAATAGTATAAATATAGCCAAAGATATCTTAAAGAAAGCAAAAGAAATAAATAATC 5774  
Qy 5358 TTCACT---AATCATAGTGGAGTATCAAAATGCTAGAAATCATGGAATAAAGCGGAGTACA 5414

Db 5773 CACATGAAAAAACTTAGTCTTTTAAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAC 5714  
Qy 5415 GCTGAATATATATCTTTGTTGACCTCGATGATGTTGTTGATAGTAGATAGTACAAAA 5474  
Db 5713 TCTCTTAATAAAGTTTTTATAGTCTCGATGATATTTGGA-ACATAAGCTTGTGAAGA 5655  
Qy 5475 TTATATTTAATATATATAAAGTAGAAGTAGATTTATCTGGT-TGTTTTAGCTACTTTT 5533  
Db 5654 GTGTATAAAAAATTTTAGATGAACAGGATGAAGTTGATTTAGTGTGTTTTCAATGCTATTGT 5595  
Qy 5534 TTCAGAAATATAAATAATTTTGAAGTGAATAATCAATATGATTTTGA 5584  
Db 5594 TCAAGATGATGTTATTTTCATATAAAAAAGTTTGACTTTTAATTTCTGGTTTTTA 5544

## RESULT 25

US-08-487-826B-13  
; Sequence 13, Application US/08487826B  
; Patent No. 5993827  
; GENERAL INFORMATION:  
; APPLICANT: Sim, Kim L.  
; APPLICANT: Chitnis, Chetan  
; APPLICANT: Miller, Louis H.  
; APPLICANT: Peterson, David S.  
; APPLICANT: Su, Kin-zhan  
; APPLICANT: Wellens, Thomas E.  
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe Martens Olson & Bear  
; STREET: 620 Newport Center Drive 16th Floor  
; CITY: Newport Beach  
; STATE: California  
; COUNTRY: US  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/487,826B  
; FILING DATE: 10-SEP-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Israelsen, Ned  
; REGISTRATION NUMBER: 29,655  
; REFERENCE/DOCKET NUMBER: NIH121.001CPI  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 235-8550  
; TELEFAX: (619) 235-0176  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 19124 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; US-08-487-826B-13  
Query Match 1.7%; Score 119.2; DB 2; Length 19124;  
Best Local Similarity 42.4%; Pred. No. 3.2e-12;  
Matches 1497; Conservative 0; Mismatches 1983; Indels 48; Gaps 14;  
Qy 2833 TCTAATATGGAATTTTAAAGATATATAAAGTGAAGAAAAATCAATCTTCTCTATTGTC 2892  
Db 3504 TGTGAATTTTTCATTTAATATGCTATGATCATTTGATTAATTAATTTTATTAATATTT 3563  
Qy 2893 TTTACAGATGATCTTATTTGATGAATTTGCTAAATGCAAGAAATTTAGGTTTTTATTGCT 2952

Db 3564 ATATTTTTTATACCTTGGATTTCTTACATTTGTTTTATTATTATATGATTTATTATTATT 3623  
Qy 2953 AGAAGTTTAAAAATAGAAAAATAAATCTAAATTTTAAAGAAATTTATTACTTAAAAATAAT 3012  
Db 3624 ATACTTATATATATATATATTTTACATTAAGATATATATATATGATCTATCTATCTATC 3683  
Qy 3013 AGTTGATTTTGTGAGAGTAATGTATGTTTAAATTTATTTAAATATATGACCCGGAATATTTTA 3072  
Db 3684 TATCTAT 3743  
Qy 3073 TTTTAAAGTACTTCTGGTTGATTTATTTTATTTCCAGAGCAAAAGTATGATTT---TTTAT 3129  
Db 3744 GTGATGATTTAATAATAAACCTTATTGAAGAGATAGAAACATAATAATATATATTTAAATTAAT 3803  
Qy 3130 TAAATTTTATGAATTTTAAATTTTATTTTATATAAAAACTTAAGCTAAATTTAA 3189  
Db 3804 AGAATTTCAATTTTATTTGTTATATATATATAAAAAATAAGAAATTTGAAAAAGTATTTTAC 3863  
Qy 3190 AAAATGAAATTTTATTTTATTTATGTTCTATATATATATGTTTGTGTTTTCAGTAGTCACAA 3249  
Db 3864 ACATGATAATGTTATTTTATTTTATTTGTTGTTTATATATTTTATATAAAAAATTTGTTTA 3923  
Qy 3250 GTATGTTTGTGAAATAAATTTTGAAGATTTATTTGCAAGATTTTACTGCTCCCATATTT 3309  
Db 3924 ATATAAGTTGTTATTAATTTTATATATGTCACCATTTAGCTTTTCCATTTATACAAATAT 3983  
Qy 3310 GGATTTTGTCAATATGTTATATATTTGTTATTTTATTAATATATGTTTATTAATAAAT 3369  
Db 3984 ATATTTCTCTCAATTA--GAATCTGAATATTTTATGTTATTAATAAAAAAGTATTAATAATA 4041  
Qy 3370 TAAAAAATAGTATCTTTTATTTAGTTTTTATTTAGTTATATCTGCAATTTATATTA 3429  
Db 4042 AAATATCTAAGATTTTCTAATTTGTTTAAATTTATATAAATTTTAAATTTATACGATA 4101  
Qy 3430 TTCAAAATGGAAAGATATGTTATTTTATGACAGA-CACCTTATAGGACTAGACTATCTT 3488  
Db 4102 GAATAAATTAATAATCAACATATATATATGTTTATCTTAAAGAACCTATTACAAATATAGT 4161  
Qy 3489 ATAACAGCGTCAAAACNAGGTTGGTTGGCTTTATGAACTATCTTACGTTAAATACCCT 3548  
Db 4162 AACAACTGGTTCTCTTTTATTAATAACATAAGAATGTGTAAAGATAGTTGTTAAA 4221  
Qy 3549 ACAATATAGTTTCAATTTCCGTTAATCTTTGCACCTTATAAAAAATAAATAAGCAAAATTT 3608  
Db 4222 GGCTTTTAAATATGATTAATAATGTTTGAAGATATATATATAGATATCTTAAACATA 4281  
Qy 3609 TTTTCTTTGCTCTTGTCTTTTATACCGATCTATTTAAGTGGATCGAGAAATTTGTTAGTTTA 3668  
Db 4282 CAACCTTTGCATAATTTGTAATTTAAAAAATAATATAATAAGAAATATTATAAATAATATT 4341  
Qy 3669 TCGGTAGCAATATTAATTAATGTTGTTTATGGAGATATATAGTGGGAAATTTGCTGG 3728  
Db 4342 ATAAAAAATTAAGCATAAATGTCACAAATAAATTTTTTTTATTAATTTTATTTTATTTTA 4401  
Qy 3729 ATAAAAAGCTAATAGTAATATTTGTAATCTACTTATTTTAAATTAAGTAAATTTGCTT 3788  
Db 4402 TTGTTCTTAAATATATTTGATTAAGAAATATTA-TTGTGCTAATATATTAATTAAGATATT 4460  
Qy 3789 TACCATGAAATTTTGGCTGTTTATATAATTTCTAGAGAAATCAAGTAACGAAAGCTAGATTATT 3848  
Db 4461 TCTAATATTAATTTATATATATATATATTTTAAAGTATTTTAAAGAAATTTTACTTATT 4520  
Qy 3849 ATTTATCA-----AGGAAGTATTGATTAAGTATTAGAAAAACAATTTTATT 3896  
Db 4521 TATTATAATAGAAATATGATGAGTATATATAAATATTGATGACAAAAAATAAATTTT 4580  
Qy 3897 GGATATGGAATATCGGATATTTTCAAGTACGGGAACTTTGGCTCGGAGTCATTTCAGGCTAT 3956  
Db 4581 TAAATGGAAATATGCAATATAAATAAATACTATATAGTATAAATTTGGTGAATAGTTGTA 4640  
Qy 3957 ATATCAATTTTTTATATAAATCAGGAATAGTTGGGTTGATTTTACTGATGTTTCTTTT 4016

Db 4641 ACTTATACAAACATGTTGCATTCATTAATTTAGAGATTATGTAAATTTGTTTATGTCATCGT 4700  
Qy 4017 TATGTTTATATAAAAGTATTAGAGTTTAAATCGGGAAACAGCATATTTTATTTTACATCA 4076  
Db 4701 AATATATATTAAATAAATGTTTTTATGATGTATGTTCTCTAAATAATATATTCATAT 4760  
Qy 4077 TTAGCCATATTTTTCATATATGAACAATAGATCCGATTTATTTATATTAGTACTATTC 4136  
Db 4761 GTAGTCATAGTGTCAATGAATATATAAATATGTTATATTTTATATTATTTTAAAT 4820  
Qy 4137 TTTTCTTCAATPAGGTATTTGGAAATAATATAAATTTTAAAGGATATAGGACACAAAAT 4196  
Db 4821 AAGTAACACAGCAATATATATAGTAATAAATAGAGAAATATATATTTTATGTTAT 4880  
Qy 4197 GAATGATTTTAAATTCAGTTATGTACCAATTTTAAATGTCCAGATTAATCTTGATAAATG 4256  
Db 4881 ATATTATTAGTTATATAAAGGGGAAATTCATATAATTTATGAAAAATTTTGTATATGA 4940  
Qy 4257 TATTAAACAGTATTATTAACCAACATATACTAATTTTAGAGGTTTATCTCGTAAATGATGG 4316  
Db 4941 TATAGTTATAGTTTAAAGGAAAAAAGGAAAAAAGGAAAAAAGGAAAAAAGT 5000  
Qy 4317 AAGTACTGATGATTTCTGAGAAAAATTCGCTTAAACTATATGAAGAACGATGGAAGAAATTA 4376  
Db 5001 TACTGTAATAGGATAAATATATATATAAATGTTTATTTTATCTTAAAAAGGTTCCCTA 5060  
Qy 4377 ATATTACAAGAAATTAATGCGGCTGACGAGATGCTCGAAATTTCCGACTAGAACATGC 4436  
Db 5061 TTATAACATTTAAAAAATTTGTCCTCATTTTATATAAATTAATTAACATATTACATAATGA 5120  
Qy 4437 AACAGGTAAATATATGCTTTTTCGATTCCTGATG-----ACTATATAGAGTTGCAATG 4491  
Db 5121 AATTCGATTTTGTGTTTTTGTAGTAATATATGGAATTAATTTATTTATGTAATGCG 5180  
Qy 4492 TTCGAGAAATGCAATGATATAATTAATCTGAGTATAATGCCATATAGCAGAGATAGATTTT 4551  
Db 5181 TTCTATATAATAATAATTTTATTTTAAAAATGAAAAATGAATAAATAATATCTCTGA 5240  
Qy 4552 TGTTTAGTACGAAAAACGGGTATACAAAGAAAAAGAAATAGTAATTTTCAATGCTCTTA 4611  
Db 5241 TTTTGTAGTCCAAATAGCTTAATAATAATTAATGACTCATATATATATATATATATCTTT 5300  
Qy 4612 ACAGAGAGAGACTGTAATAAGAAATTTTGTGCAGGATC-----TAATATAGA 4658  
Db 5301 ACAACAATATAAGTAAATATTTTAACTTTTAAAGGAAAAATAAATAAATAAAT 5360  
Qy 4659 AAATAATGTTTGGTGCAGCTTTTATCCAGAGATTTATATAAGATATAAATAATTTCCAAAT 4718  
Db 5361 AAGAATACTGAATAATAGTCATATTATACATTTTAAAAATGTAACATAAATTACAAAT 5420  
Qy 4719 TAATAATAGAGTATTCGGTAGGATTTGCTTTTAAATTT--GGAGGCTCTTGAAACAATGTA 4776  
Db 5421 ACGTAAACATGTTATATAGAAATAATAAGAAATTTAATAATTAAGGATAAATAATAATTTA 5480  
Qy 4777 ACAGCTGTAGTTGATACCTAGAGATATTTATTAATTTATGTCATTCGTACAGTTGCG 4836  
Db 5481 AAATTAATTTTTTATGTCATTTTATGTTATATATATATATAATTAACATGATTAGTTTT 5540  
Qy 4837 CTTAATTAATCAGAAATTTCTTATATAATAATAATTCATTTAGTTCACAGATTGGAAATTAC 4896  
Db 5541 TTGAAAAATTTTAAATCATATAATAATAATAATAATAATTTAGTTAAATAATAATAGTATTTCA 5600  
Qy 4897 CCCTTTAAGTTTAAAAAGAGGTTTATGTCATTTTATGATGCAAAAGTTATTAAGAGAG 4956  
Db 5601 ACAAAATACTAACTTATAAGTATATCATATAAATATATATATATATATATATTTATGTTT 5660  
Qy 4957 GTTAAATGTTTAAACAAAATGTTTCAACAGATTTGTTTGGATATGATGTTCTCCCAATA 5016  
Db 5661 TTGATGGGTGATATATAGGCTATAAGGTATATATATATGTTGTTTCAATATATATATATATG 5720  
Qy 5017 TTAGAGCTTTATCGAAAGAAATACGTAGATATCCATTTTATTAAGCGCAAAAGATATTTA 5076  
Db 5721 GAATAGATACATAAAGTT--AATATATTTTATGTTATATGCTGCTGTTAAGATAGATA 5779

Qy 5077 T-CAAGAAAGCATTTTAGTTACGTTGTATTTTGTAGTAAATTTTCGCTTAAACTATATGTAAT 5135  
Db 5780 TGCATTTACAGTTAAGGTTTATAGTTTTTTTTTTTTTTTTTTTTTTTGTACATATATAAAAAAT 5839  
Qy 5136 GTTATATAAGAAATTTCAAAAGCAGTAGAGGTAAAAATGATAAAATTTAGTGTATTGTT 5195  
Db 5840 AGATACTAACATATGTCATATTACAGATAAATATTTGTATAAATAATATATATATATA 5899  
Qy 5196 CCAGTTTATAATGTAGATAAATAATTTAAGTAGTTGTATAGAAAGCAATTAATAATCAAAAT 5255  
Db 5900 TATATATAAA--GACATTAAAACTATACTAATAGGTAATTTAGTTTTTATATATCATCCT 5956  
Qy 5256 TATAAAAAATATAGAAATATTTTGTATAGATGATGGCTCTGTAGATGATTTCTGCTAAATA 5315  
Db 5957 TTTATTATTATAAATTTTTTTTGTGTTTACTCTTCTGCTGTTTTTTTGTGTTTATATAATA 6016  
Qy 5316 TGCAGGAATATGCGAGAAAGATAAAAGAGTAAAAATTTTTTTCACATAATCATAGTGA 5375  
Db 6017 ACAATATATAACAATATCAGTATTTGGAAATATAAATAAATTTATTTCTACATATGCAAT 6076  
Qy 5376 GTATCAAAATGCTAGAAATCATGGAATAAAGCGGAGTACAGCTGAATATATATTGTTGTT 5435  
Db 6077 AT 6135  
Qy 5436 GACTCTGATGATGTTGTTGATAGTAGATAGTAGTAGTAAATAATATATTTTAATATATATAA 5495  
Db 6136 TACTATTTTATATACATGCAATTTTATATATATTTTATATATATTTTAAAGATATATTTAAT 6195  
Qy 5496 AGTAGAAGTATTTATCTGTTGTTGTTGACGCTACTTTTTTCAGAAATAATAAATAATTTT 5555  
Db 6196 ATTTATATAGTACATATATGTTATTTATATATAAACAATTTTCAATTTATATATAAATAT 6255  
Qy 5556 GAAAGTGAATAATCAAAATATTTGATTTTGAAGCAATTAATCCGTGCGAGCATGGGAGAA 5615  
Db 6256 ATAGAACATGAACATTTTATTAATAACTCATATTTGAAATATATATTTTATATATGCTAT 6315  
Qy 5616 AAAAAATTTTATGAATTTGTATATAAATAATATTTTTTTCTACTCT--GTTGTAAACTATA 5674  
Db 6316 TTTTACTTTATTTTATATATATACAAATAAAATTTTGAATTTTCAATAAATGCAATGAATA 6375  
Qy 5675 TAAGAAAAGATACATAACAGATCTTTTCAAGAGAAATCAATGGTTAGGAGAGATTTTACT 5734  
Db 6376 CATAAAAAATAACAAACAATAATGATAAACAATTTTATTAATAATAATAATAATAATAA 6435  
Qy 5735 TTTTAATCTGCATTTTAAAGAAATATAGATAGAGTTAGTTATTTTGAAGCAATCTTTA 5794  
Db 6436 TATAATAATATTTTTTCCGTGTTATTTATTTATCAATTTTTTTTTTTGATGCTATATATAT 6495  
Qy 5795 TTTTATAGGAGAGGTATCTAAGTACAGTAAATTTCTTTTAAAGAGGTGTTTTTGGCA 5854  
Db 6496 ATTATATAAATAATTAATAATAATAACAACAAAAATTAATAATAATAATAATAATAATAATA 6555  
Qy 5855 ATTGGAAAAATTTTCAAAAAACAAGTATGTTTGTAAAGCAATAATATATGCTGAGGATTT 5914  
Db 6556 AATATATACAAATFACAAAGAAATGTTATCTATATCAATTTATATATATATATAGATATAT 6615  
Qy 5915 TGAAGTATCAATTTGTTAAAGATATATACGTTGGCAAGTATTTTATTTATAGCTTACTAAT 5974  
Db 6616 AATATATAGATAAATAATAGATAGAGAGAAACGAAGACATAATTTGCTCTCTTTGTTATC 6675  
Qy 5975 GTTTAAATACGGAACACAGTCTATTTTGTACAAATTTTAAATTTTGTAGAAATCTTTATAA 6034  
Db 6676 TCTAATATATATATATATAATAAATAAATAAAGTCAAAAAATAATACATATATTA 6735  
Qy 6035 ABAATATATTTTAACTGTTTAAAGTATCTACAAAAATTTCTTGTCTAAAAATTTTGTG 6094  
Db 6736 AGTTTAATTAATAATATAAACAACGTTGCAATATATCTTTTATATATGTTTGTATTTT 6795  
Qy 6095 TATAAGAAATGTTTCGAACAAAGTTTTTAAAAAATAATTTATGTTTAAATAGGAAGATAT 6154  
Db 6796 CGTATTTTTTTTTCTCATTTATAATTTTACTTAAATAAATAAACAATAAAAAATAATA 6855

Qy	6155	CATGGATACT - ATTAGTAAATTTCTATAAATTGTACCTATATATAATGTAGAAAAATAT	6212
Db	6856	TATATATAAATTAAATAGATAAAAGGATACATAAAATATAATTTCTGATTATATT	6915
Qy	6213	TTATCTAAATGCTATAGTAGCATTTGTAATCAGACCTACAAACATATAGAGATTCTCTG	6272
Db	6916	TTTTTTTGTAGAAATTTTAAATTTATTATAAATTTTAAATATATATATATATATTTTTT	6975
Qy	6273	GTGAATCACGCTAGTACGATAATTTCGGAAGAAAATTTGTTTAGCATAT	6320
Db	6976	TTAAAAATATATAAAACTAATAATTATTATTATATACATATTAAATAT	7023

Search completed: April 28, 2005, 15:47:29  
Job time : 1118 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 28, 2005, 14:17:37 ; Search time 3571 Seconds  
(without alignments)  
11919.659 Million cell updates/sec

Title: US-09-767-041-9  
Perfect score: 6992  
Sequence: 1 atcgcaaacgaattggca.....gcaaacagcaaaagagctt 6992

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 5642217 seqs, 3043843248 residues

Total number of hits satisfying chosen parameters: 11284434

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/PCT NEW PUB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT NEW PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06 NEW PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06 PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07 NEW PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08 NEW PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08 PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09 NEW PUB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10C PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10D PUBCOMB.seq.\*
- 17: /cgn2\_6/ptodata/1/pubpna/US10E PUBCOMB.seq.\*
- 18: /cgn2\_6/ptodata/1/pubpna/US10F PUBCOMB.seq.\*
- 19: /cgn2\_6/ptodata/1/pubpna/US10 NEW PUB.seq.\*
- 20: /cgn2\_6/ptodata/1/pubpna/US11 NEW PUB.seq.\*
- 21: /cgn2\_6/ptodata/1/pubpna/US60 NEW PUB.seq.\*
- 22: /cgn2\_6/ptodata/1/pubpna/US60 PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	6992	100.0	6992	9	US-09-767-041-9
2	6992	100.0	6992	9	US-09-767-041-29
3	727.8	10.4	6865	9	US-09-900-038A-3
4	727.8	10.4	6865	19	US-10-865-873-3
5	726.8	10.4	17276	9	US-09-870-759-83
6	726.8	10.4	17276	10	US-09-751-708A-83
7	726.8	10.4	17276	18	US-10-428-817A-79
8	721.4	10.3	25020	17	US-10-192-280-1
9	684.8	9.8	18373	17	US-10-461-990-4
10	652.4	9.3	2226	18	US-10-804-408-166
11	652.4	9.3	2226	18	US-10-804-408-167

12	650.8	9.3	2217	18	US-10-804-408-162	Sequence 162, App
13	649.2	9.3	2226	18	US-10-804-408-171	Sequence 173, App
14	648.8	9.3	2226	18	US-10-804-408-171	Sequence 171, App
15	647.6	9.3	2217	18	US-10-804-408-172	Sequence 172, App
16	647.6	9.3	2226	18	US-10-804-408-168	Sequence 168, App
17	647.6	9.3	2226	18	US-10-804-408-169	Sequence 169, App
18	646	9.2	2217	18	US-10-804-408-164	Sequence 164, App
19	644.4	9.2	2217	18	US-10-804-408-163	Sequence 163, App
20	644.4	9.2	2226	18	US-10-804-408-170	Sequence 170, App
21	638.8	9.1	2225	18	US-10-804-408-165	Sequence 165, App
22	428.4	6.1	6850	16	US-10-182-960-1	Sequence 1, Appli
23	423.2	6.1	1368	9	US-09-815-242-9587	Sequence 9587, Ap
24	419.8	6.0	2581	18	US-10-804-408-178	Sequence 178, App
25	386	5.5	2577	18	US-10-804-408-179	Sequence 179, App
26	336.6	4.8	2722	18	US-10-804-408-176	Sequence 176, App
27	318.4	4.6	2692	18	US-10-804-408-177	Sequence 177, App
28	301.4	4.3	8056	18	US-10-473-126-386	Sequence 386, App
29	298.8	4.3	8056	18	US-10-473-126-386	Sequence 386, App
30	222.2	3.2	8056	18	US-10-473-126-240	Sequence 240, App
31	219.8	3.1	8056	18	US-10-473-126-240	Sequence 240, App
32	206.6	3.0	50000	19	US-10-706-635-23	Sequence 23, Appli
33	205.2	2.9	3673778	16	US-10-312-841-1	Sequence 1, Appli
34	195.2	2.8	3673778	16	US-10-312-841-1	Sequence 1, Appli
35	187.8	2.7	5979	14	US-10-239-676-18	Sequence 18, Appli
36	187.8	2.7	5979	15	US-10-240-453-26	Sequence 26, Appli
37	185.8	2.7	50000	19	US-10-706-635-23	Sequence 23, Appli
38	183.6	2.6	8170	15	US-10-240-453-132	Sequence 132, App
39	179	2.6	11691	15	US-10-311-455-2214	Sequence 2214, Ap
40	176	2.5	372	17	US-10-282-122A-38093	Sequence 38093, A
41	176	2.5	2337	18	US-10-804-408-175	Sequence 175, App
42	174.6	2.5	3673778	16	US-10-312-841-2	Sequence 2, Appli
43	172.8	2.5	15373	15	US-10-311-455-440	Sequence 440, App
44	172.8	2.5	19634	18	US-10-473-126-302	Sequence 302, App
45	171.4	2.5	2384	18	US-10-804-408-174	Sequence 174, App

ALIGNMENTS

RESULT 1  
US-09-767-041-9  
; Sequence 9, Application US/09767041  
; Patent No. US20020055168A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Hilda  
; TITLE OF INVENTION: STREPTOCOCCUS SUIIS VACCINES AND DIAGNOSTIC TESTS  
; FILE REFERENCE: 2183-4726  
; CURRENT APPLICATION NUMBER: US/09/767,041  
; CURRENT FILING DATE: 2001-01-22  
; PRIOR APPLICATION NUMBER: PCT/NL99/00460  
; PRIOR FILING DATE: 1999-07-19  
; PRIOR APPLICATION NUMBER: EP98202465.5  
; PRIOR FILING DATE: 1998-07-22  
; PRIOR APPLICATION NUMBER: EP98202467.1  
; PRIOR FILING DATE: 1998-07-22  
; NUMBER OF SEQ ID NOS: 53  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 9  
; LENGTH: 6992  
; TYPE: DNA  
; ORGANISM: Streptococcus suis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(6992)  
; OTHER INFORMATION: CPS 2  
US-09-767-041-9

Query Match 100.0%; Score 6992; DB 9; Length 6992;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 6992; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Cyt 1 ATCCCAACGAATTCGCAATTATTCATATGATGACAGTTGCAATTCCTCAATCTTAA 60  
|||||

Db 1 ATGCCAAAGCAAATTCGCATTAATTTGATGATAGCAGTTGCAATTTCTGCAATCTTAA 60  
Qy 61 CAAGTCATATACCAAATGCTGATTTAAATCGTTCTGGAAATTTTATCATTAATGATGTTTC 120  
Db 61 CAAGTCATATACCAAATGCTGATTTAAATCGTTCTGGAAATTTTATCATTAATGATGTTTC 120  
Qy 121 ATTATTTTGGCAATTTTATATCTCGTATGCCAGTTGGAATTTGAGTATAGAGGTAATCTGA 180  
Db 121 ATTATTTTGGCAATTTTATATCTCGTATGCCAGTTGGAATTTGAGTATAGAGGTAATCTGA 180  
Qy 181 TAGAGTTTGGAAAAACAATTTAACTATAGTATAATTTTGGCAATTTTCTTACGGCAGTAT 240  
Db 181 TAGAGTTTGGAAAAACAATTTAACTATAGTATAATTTTGGCAATTTTCTTACGGCAGTAT 240  
Qy 241 CATTTTGTGGGAATAATTTCCGACCTTTCAAGACGTCGTGCGGTGTAATTTTCACATTAA 300  
Db 241 CATTTTGTGGGAATAATTTCCGACCTTTCAAGACGTCGTGCGGTGTAATTTTCACATTAA 300  
Qy 301 TAAACTTCGTTTGGTATACCTATTTAACTGTAATTTAAGCAGTTTAAAGATAGCTTTC 360  
Db 301 TAAACTTCGTTTGGTATACCTATTTAACTGTAATTTAAGCAGTTTAAAGATAGCTTTC 360  
Qy 361 TATTTTCCGACAACTATCAAAAAAGACGATTTCTAAATTAACAACGCTGAACGATGGGAAA 420  
Db 361 TATTTTCCGACAACTATCAAAAAAGACGATTTCTAAATTAACAACGCTGAACGATGGGAAA 420  
Qy 421 ATATGCAAGTTTATTTGAATFACATAAACAAATTCAAAAAATCTTTGTTGCAATGGTAG 480  
Db 421 ATATGCAAGTTTATTTGAATFACATAAACAAATTCAAAAAATCTTTGTTGCAATGGTAG 480  
Qy 481 TTTTAGGTACAGAAATAGATAAAATTAATTTATCAATTTACCGCTCTATTTCTGTGGAAG 540  
Db 481 TTTTAGGTACAGAAATAGATAAAATTAATTTATCAATTTACCGCTCTATTTCTGTGGAAG 540  
Qy 541 AAGCTATAGAGTTTTCACAAAGGAAGTGGTCGACACAGCTCTTTATATAATCTACCAAGTG 600  
Db 541 AAGCTATAGAGTTTTCACAAAGGAAGTGGTCGACACAGCTCTTTATATAATCTACCAAGTG 600  
Qy 601 AGTTTTTAGACGTAAGCAATTCGTTTTCAGATTTTGAAGTTGTTAGGTAATGATGAAGCG 660  
Db 601 AGTTTTTAGACGTAAGCAATTCGTTTTCAGATTTTGAAGTTGTTAGGTAATGATGAAGCG 660  
Qy 661 TTGATATTAATTCATTCGGTTTACTCGGTTTGAAGAAACAAAAATCCAACTGCTAGGTG 720  
Db 661 TTGATATTAATTCATTCGGTTTACTCGGTTTGAAGAAACAAAAATCCAACTGCTAGGTG 720  
Qy 721 ACCATAGCATTTGAATCTTTTCCAAATTTTATAAGCCCTAGTCATATCATGATGAAC 780  
Db 721 ACCATAGCATTTGAATCTTTTCCAAATTTTATAAGCCCTAGTCATATCATGATGAAC 780  
Qy 781 GACTTTTGGATATACTCGGAGCGTAGTCGGGTTAAATTTTGGTATAGTTTCTATTT 840  
Db 781 GACTTTTGGATATACTCGGAGCGTAGTCGGGTTAAATTTTGGTATAGTTTCTATTT 840  
Qy 841 TGTTAGTTCCAAATTTTCTAGAGATGGTCGACCGCTATTTTGGCTCAGAAACGAGTTG 900  
Db 841 TGTTAGTTCCAAATTTTCTAGAGATGGTCGACCGCTATTTTGGCTCAGAAACGAGTTG 900  
Qy 901 GACAGAATGGACGATATTTACATTTCTACAGTTTTCGATCGATGATGTTGATGCTCAGG 960  
Db 901 GACAGAATGGACGATATTTACATTTCTACAGTTTTCGATCGATGATGTTGATGCTCAGG 960  
Qy 961 AGCGCAAAAAAGACTTGTCTAGCCAAAAACAGATGCAAGGGTGGGTATGTTTTAAAAATGG 1020  
Db 961 AGCGCAAAAAAGACTTGTCTAGCCAAAAACAGATGCAAGGGTGGGTATGTTTTAAAAATGG 1020  
Qy 1021 GAAAAACGATCCTAGAAATTTACTCCAATTTGGACATTTTATAGCGAAAAACAAAGTTTACAG 1080  
Db 1021 GAAAAACGATCCTAGAAATTTACTCCAATTTGGACATTTTATAGCGAAAAACAAAGTTTACAG 1080  
Qy 1081 AGTTTACCACAGTTTATTAATGTTTTTAAATGSCGATATGAGTCTAGTTGTTGACGCTCCAC 1140  
Db 1081 AGTTTACCACAGTTTATTAATGTTTTTAAATGSCGATATGAGTCTAGTTGTTGACGCTCCAC 1140

Qy 1141 CTRACAGTTGATGAAATTTGAAAAATATATCTCCTGGTCAAAAGAGACGATTCGAGTTTAAAC 1200  
Db 1141 CTRACAGTTGATGAAATTTGAAAAATATATCTCCTGGTCAAAAGAGACGATTCGAGTTTAAAC 1200  
Qy 1201 CAGGGATTAACAGTCTCTGCGAGTTAGTGGTCTAGTATATATACACAGACTTCGACGACG 1260  
Db 1201 CAGGGATTAACAGTCTCTGCGAGTTAGTGGTCTAGTATATATACACAGACTTCGACGACG 1260  
Qy 1261 TAGTTCGGTTGGACTTTAGCATACATTTGATAAATTTGGACTCTCTGGTCTAGATATTTAAATTT 1320  
Db 1261 TAGTTCGGTTGGACTTTAGCATACATTTGATAAATTTGGACTCTCTGGTCTAGATATTTAAATTT 1320  
Qy 1321 TATTAAGACAGTGAAGTTGTTATTTGTTGAGAGAGGGAAGTAAGTATATGAAAG 1380  
Db 1321 TATTAAGACAGTGAAGTTGTTATTTGTTGAGAGAGGGAAGTAAGTATATGAAAG 1380  
Qy 1381 TTTGTTTGGTTCGGTCTCTCAGGGGACATTTGACTCACTTGTATTTGTTAAAAACCGTTT 1440  
Db 1381 TTTGTTTGGTTCGGTCTCTCAGGGGACATTTGACTCACTTGTATTTGTTAAAAACCGTTT 1440  
Qy 1441 GGAAGGAAGAAACGTTTTTGGGTAAACATTTGATAAAGAGGATGCAAGAGTCTTTTGA 1500  
Db 1441 GGAAGGAAGAAACGTTTTTGGGTAAACATTTGATAAAGAGGATGCAAGAGTCTTTTGA 1500  
Qy 1501 AGAATGAAAAAATGTATCCATGTTACTTTCCAAACAAATCGCAATCTCAATTAATTTAGTGA 1560  
Db 1501 AGAATGAAAAAATGTATCCATGTTACTTTCCAAACAAATCGCAATCTCAATTAATTTAGTGA 1560  
Qy 1561 AAAATACTTCTTAGCTCTTCAAAATTTTACGTGATGAGAAACAGATGTTTATTTTTCAT 1620  
Db 1561 AAAATACTTCTTAGCTCTTCAAAATTTTACGTGATGAGAAACAGATGTTTATTTTTCAT 1620  
Qy 1621 CTGGTCGGCCGTTGCTGCTCCCTTTTACATTCGAAAAATATTTTGGAGCAAGACGA 1680  
Db 1621 CTGGTCGGCCGTTGCTGCTCCCTTTTACATTCGAAAAATATTTTGGAGCAAGACGA 1680  
Qy 1681 TTTATATTTGAAGTATTTGATTCGAGTTTAAATAATCTACATTTAACTGGAATACTAGTTTATC 1740  
Db 1681 TTTATATTTGAAGTATTTGATTCGAGTTTAAATAATCTACATTTAACTGGAATACTAGTTTATC 1740  
Qy 1741 CCGTAACAGATTTTTTATTTGTTTCACTGGGAAGAAATGAAGGTATATCTTAAATCTA 1800  
Db 1741 CCGTAACAGATTTTTTATTTGTTTCACTGGGAAGAAATGAAGGTATATCTTAAATCTA 1800  
Qy 1801 TTAACCTTGGGAGTATTTTTTAAATGATTTTGTAAACAGTAGGAACTCATCAACAAACAGTT 1860  
Db 1801 TTAACCTTGGGAGTATTTTTTAAATGATTTTGTAAACAGTAGGAACTCATCAACAAACAGTT 1860  
Qy 1861 TAATCGATTTGATAAAAGAGATTCATTTATTTGAAAAAATAATGGAAGTATAACCGACGAAT 1920  
Db 1861 TAATCGATTTGATAAAAGAGATTCATTTATTTGAAAAAATAATGGAAGTATAACCGACGAAT 1920  
Qy 1921 ATTTATTTCAACACAGGATTTCTGACTATATTTCCAGAAATTTGCAAGTATATAAATAATTTCT 1980  
Db 1921 ATTTATTTCAACACAGGATTTCTGACTATATTTCCAGAAATTTGCAAGTATATAAATAATTTCT 1980  
Qy 1981 CAGTTACAAAGAAATGGAACAAATATTAACAATCAGAAATGAGTTATTTTCCACCGAGG 2040  
Db 1981 CAGTTACAAAGAAATGGAACAAATATTAACAATCAGAAATGAGTTATTTTCCACCGAGG 2040  
Qy 2041 CCCCCTACTTTTATGAAATTTCAATTTCCAAAGAAAAAACAATTTATTTTCTAGACA 2100  
Db 2041 CCCCCTACTTTTATGAAATTTCAATTTCCAAAGAAAAAACAATTTATTTTCTAGACA 2100  
Qy 2101 AAAAAAGTATGGTGAACATGTAATCATCAAGTAGAGTTTGTAAAGAAATTTTAC 2160  
Db 2101 AAAAAAGTATGGTGAACATGTAATCATCAAGTAGAGTTTGTAAAGAAATTTTAC 2160  
Qy 2161 AGATAATAATTTTATTTATAGAAAAATAGATGATTTGTTTGAATAAATTTATTTGAAGT 2220  
Db 2161 AGATAATAATTTTATTTATAGAAAAATAGATGATTTGTTTGAATAAATTTATTTGAAGT 2220

QY	2221	TTCTAAGCAACTAACCTTTACATCAATAATAATATTTTGTGGAAGATTAAACAAAT	2280
DB	2221		2280
QY	2281	TTCTAAGCAACTAACCTTTACATCAATAATAATATTTTGTGGAAGATTAAACAAAT	2340
DB	2281		2340
QY	2341	AGTTGAAAAATTTAATGAGATCAAGAAATGAATAAATAAAGATGCATATTTGATAA	2400
DB	2341		2400
QY	2401	AGTTGAAAAATTTAATGAGATCAAGAAATGAATAAATAAAGATGCATATTTGATAA	2460
DB	2401		2460
QY	2461	TGCTTTATCATATTTTCTCAGATTTTCTCGAGAGGATACAGATATTTATCATCTTCT	2520
DB	2461		2520
QY	2521	TGCTTTATCATATTTTCTCAGATTTTCTCGAGAGGATACAGATATTTATCATCTTCT	2580
DB	2521		2580
QY	2581	CTCAGAGATGACACACCAATAGTTCTCAGATTTTCTCGAGAGGATACAGATATTTATCATCTTCT	2640
DB	2581		2640
QY	2641	CTCAGAGATGACACACCAATAGTTCTCAGATTTTCTCGAGAGGATACAGATATTTATCATCTTCT	2700
DB	2641		2700
QY	2701	CTCAGAGATGACACACCAATAGTTCTCAGATTTTCTCGAGAGGATACAGATATTTATCATCTTCT	2760
DB	2701		2760
QY	2761	AGTAACTAAATGTCAGATGAACTATTATATACAGCAATATAGAAATATGAAATTT	2820
DB	2761		2820
QY	2821	AGTAACTAAATGTCAGATGAACTATTATATACAGCAATATAGAAATATGAAATTT	2880
DB	2821		2880
QY	2881	TCCTCTATTGCTTTACAGATGAACTATTATATACAGCAATATAGAAATATGAAATTT	2940
DB	2881		2940
QY	2941	TCCTCTATTGCTTTACAGATGAACTATTATATACAGCAATATAGAAATATGAAATTT	3000
DB	2941		3000
QY	3001	AAAAATATAATGTCAGATGAACTATTATATACAGCAATATAGAAATATGAAATTT	3060
DB	3001		3060
QY	3061	CGAATATTTTATTTTAACTCTCTGTTGATTTTATTTTATTTTATTTTATTTTATTTT	3120
DB	3061		3120
QY	3121	TATTTTATTTTATTTTAACTCTCTGTTGATTTTATTTTATTTTATTTTATTTTATTTT	3180
DB	3121		3180
QY	3181	TAAATATAAATAAGAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT	3240
DB	3181		3240
QY	3241	TAGTCACAGTATGTTTGTGAAATATAAATTTTGAAGATTTTTCAGATTTTACTGCTC	3300
DB	3241		3300
QY	3301	CCATAATTTGCAATTAATGATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTT	3360
DB	3301		3360

DB	3301	CCATAATTTGCAATTAATGATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTT	3360
QY	3361	ATAAAAATTTAAAAATAGTATCTTTTATTTTATTTTATTTTATTTTATTTTATTTT	3420
DB	3361		3420
QY	3421	TGTATATTTTCAAAATGGAAGATATTTGATTTTATTTTATTTTATTTTATTTTATTTT	3480
DB	3421		3480
QY	3481	ACTATCTTATAACAGCGCTCAAAAACAAAGTTGCTTGTGCTTTATGAACTATCTTACGTTAA	3540
DB	3481		3540
QY	3541	ATACCACTACAATTTATGATTTTCAATTTCCGTTTAACTTTTCACTTATAAAAAATAAAAATGC	3600
DB	3541		3600
QY	3601	AACAATTTTTTTTCTGCTGCTTTTATACCGATCTATTTAAAGTGGATCGAGAAATG	3660
DB	3601		3660
QY	3661	GTAGTTTATCGCTAGCAATATTAATTTATTTATTTATTTATTTATTTATTTATTTATTT	3720
DB	3661		3720
QY	3721	TTGCTTGGATAAAAAGCTAAATAGTAAATTTTGTAACTATCTTATTTTAAATACTG	3780
DB	3721		3780
QY	3781	AAATGCTTTACCATGAAATTTTGGCTGTTTAAATTTCTAGAGATATAGGTGGAAAT	3840
DB	3781		3840
QY	3841	GAATTTATTTTATCAAGAGATTTTGAATAAGTATTTAGAAAACAATTTTATTTTGGAT	3900
DB	3841		3900
QY	3901	ATGGAATATCCGAATATTTCAAGTATTTGCTTGGCTCGGAAGTCAATTTTATTTTAT	3960
DB	3901		3960
QY	3961	CAATTTTATTTTATCAAGATTTTGGTGTGATTTTATCTGATTTTCTTTTTTATG	4020
DB	3961		4020
QY	4021	TTATATAAAGATTTATGAGTTTAAATGGAAGATTTGGAAGATTTTATTTTATTTTAT	4080
DB	4021		4080
QY	4081	CCATATTTTTCATATATGAAACAATAGATCCGATTTATTTATTTATTTATTTTATTTT	4140
DB	4081		4140
QY	4141	CTTCAATAGTATTTTGAATATATATAATTTTAAAAAGGATATGAGACAAAAATGAAT	4200
DB	4141		4200
QY	4201	GAATTTATTTTCAATTTATGTAACCAATTTATATGTCGAAGTATCTTGATTAATGATTT	4260
DB	4201		4260
QY	4261	AACAGTATTTTAAACCAACATATCTAATTTATGAGTTTATCTTCTGATTAATGATGAGT	4320
DB	4261		4320
QY	4321	ACTGATGATTTCTGAGAAAAATTTGCTTAACTATATGAAGATCGATGGAAGATTTAAATAT	4380
DB	4321		4380
QY	4381	TACAAGAAAAATTAATCGCGGTCTAGCAGATGCTCGAAATTTTCGACTAGAACATGCAACA	4440
DB	4381		4440



4381 TACAAGAAAAATAATGGCGGTCTAGCAGATGCTCGAAATTTCCGGACTAGAAATGCAACA 4440  
4441 GGTAAATATATGCTTTTGTGCGATTCCTGATGACTATATAGAAAGTTGCCAATGTTCCAGAGA 4500  
4441 GGTAAATATATGCTTTTGTGCGATTCCTGATGACTATATAGAAAGTTGCCAATGTTCCAGAGA 4500  
4501 ATGCATGATAATATAACTCAGTATATAATCGCGATATAGCAGATAGATATTTTGTGTTAGTA 4560  
4501 ATGCATGATAATATAACTCAGTATATAATCGCGATATAGCAGATAGATATTTTGTGTTAGTA 4560  
4561 GACGAAAAACGGGATACAAAGAAAAAGAAATAGTAAATTTTTCATGCTCTTAACGAGAGAA 4620  
4561 GACGAAAAACGGGATACAAAGAAAAAGAAATAGTAAATTTTTCATGCTCTTAACGAGAGAA 4620  
4621 GAGACTGTAAGAGAAATTTTGTGCGATCTAATATAGAAAAATAAATGTTGGTGGCAGCTT 4680  
4621 GAGACTGTAAGAGAAATTTTGTGCGATCTAATATAGAAAAATAAATGTTGGTGGCAGCTT 4680  
4681 TATTACAGAGATATATAAAGATATAAAATTCCAAATTAATAAGAAAGTATTCGTGAG 4740  
4681 TATTACAGAGATATATAAAGATATAAAATTCCAAATTAATAAGAAAGTATTCGTGAG 4740  
4741 GATTTGCTTTTAAATTTGGAGTCTTTGAACAATGTGAACACGTTAGTACTAGTACTAGA 4800  
4741 GATTTGCTTTTAAATTTGGAGTCTTTGAACAATGTGAACACGTTAGTACTAGTACTAGA 4800  
4801 GAATATATATTAATGATGCTGTAACAGTTCGCTTATTAATCAGAAAAATTCCTATA 4860  
4801 GAATATATTAATGATGCTGTAACAGTTCGCTTATTAATCAGAAAAATTCCTATA 4860  
4861 AATAATATTTGATGCTCAAGATGGAAGATTAACCCCTTTAAAGTTAAAAAGAGAGTTT 4920  
4861 AATAATATTTGATGCTCAAGATGGAAGATTAACCCCTTTAAAGTTAAAAAGAGAGTTT 4920  
4921 AGTCATTATTTGATGCTCAAGATGGAAGATTAACCCCTTTAAAGTTAAAAAGAGAGTTT 4980  
4921 AGTCATTATTTGATGCTCAAGATGGAAGATTAACCCCTTTAAAGTTAAAAAGAGAGTTT 4980  
4981 TCACAGATGTTTGGATTAATGAGTTCTTGGCCAAATATAGAGTCTTTATCGAAAAAGAAATA 5040  
4981 TCACAGATGTTTGGATTAATGAGTTCTTGGCCAAATATAGAGTCTTTATCGAAAAAGAAATA 5040  
5041 CGTAGATATCCATTTATTAAGCGAAAGATATTTATCAAGAACATTTTACGTTG 5100  
5041 CGTAGATATCCATTTATTAAGCGAAAGATATTTATCAAGAACATTTTACGTTG 5100  
5101 TATTTGATGAAATTTTCGCTAAACTATATGTAATGTAATAGAGTCTTTATCGAAAAAGAAATA 5160  
5101 TATTTGATGAAATTTTCGCTAAACTATATGTAATGTAATAGAGTCTTTATCGAAAAAGAAATA 5160  
5161 TAGAGTAAATGGAATAAATTTAGTGTATTTGTTCCAGTTTATATGTAATAGTAATAATATT 5220  
5161 TAGAGTAAATGGAATAAATTTAGTGTATTTGTTCCAGTTTATATGTAATAGTAATAATATT 5220  
5221 TAAGTAGTTGTATAGAAAGCATTTATTAATCAAAATTTATAGAAATATATTATGCA 5280  
5221 TAAGTAGTTGTATAGAAAGCATTTATTAATCAAAATTTATAGAAATATATTATGCA 5280  
5281 TAGATGATGGCTCTGTAGATGATTTCTGCTAAATATGCAAGAAATGTCAGAAAAAGATA 5340  
5281 TAGATGATGGCTCTGTAGATGATTTCTGCTAAATATGCAAGAAATGTCAGAAAAAGATA 5340  
5341 AAAGAGTAAATTTTTCATCAATATAGTGGAGTATCAAAATGCTAGAAATCATGGAA 5400  
5341 AAAGAGTAAATTTTTCATCAATATAGTGGAGTATCAAAATGCTAGAAATCATGGAA 5400  
5401 TAAAGCGAGTACAGCTGAATATATGTTGTTGACCTGATGATGTTGTTGATAGTA 5460  
5401 TAAAGCGAGTACAGCTGAATATATGTTGTTGACCTGATGATGTTGTTGATAGTA 5460  
5461 GATTAGTAGAAAAATTTATATTTTAAATATATAAAAAGTAGAAGTATCTGTTGTT 5520  
5461 GATTAGTAGAAAAATTTATATTTTAAATATATAAAAAGTAGAAGTATCTGTTGTT 5520  
5461 GATTAGTAGAAAAATTTATATTTTAAATATATAAAAAGTAGAAGTATCTGTTGTT 5520

QY 5521 TGTAGCTACTTTTTCAGAAAAATATAAATAATTTTGAAGTGAATAATCCAAATATTGATT 5580  
Db 5521 TGTAGCTACTTTTTCAGAAAAATATAAATAATTTTGAAGTGAATAATCCAAATATTGATT 5580  
QY 5581 TTGAAGCAATTAATACCGTGCAGGACATGGGAGAAAAAATTTTATGAATTTGTATATAA 5640  
Db 5581 TTGAAGCAATTAATACCGTGCAGGACATGGGAGAAAAAATTTTATGAATTTGTATATAA 5640  
QY 5641 ATAATATTTTCTACTCTCTGTTTGTAAACTATATAAGAAAAAGATACATAACAGATCTTT 5700  
Db 5641 ATAATATTTTCTACTCTCTGTTTGTAAACTATATAAGAAAAAGATACATAACAGATCTTT 5700  
QY 5701 TTCAAGAGAAATCAATGGTTAGGAGAAAGATTTTACCTTTTAACTCTGCATTTATTTAAAGAAATA 5760  
Db 5701 TTCAAGAGAAATCAATGGTTAGGAGAAAGATTTTACCTTTTAACTCTGCATTTATTTAAAGAAATA 5760  
QY 5761 TAGATAGAGTTAGTTATTTGACTGAACATCTTTATTTTTATAGAGAGGTATACCTAAAGTA 5820  
Db 5761 TAGATAGAGTTAGTTATTTGACTGAACATCTTTATTTTTATAGAGAGGTATACCTAAAGTA 5820  
QY 5821 CAGTAAATTTCTTTTAAAGAGAGTGTGTTTTTGCATTTGCAAAATTTGCAAAAACCAAGTGA 5880  
Db 5821 CAGTAAATTTCTTTTAAAGAGAGTGTGTTTTTGCATTTGCAAAATTTGCAAAAACCAAGTGA 5880  
QY 5881 TAGTATGTTTTAAGCAAAATATATGTTGAGGATTTTGCAGTATCAATTTGTTAAAGATACTA 5940  
Db 5881 TAGTATGTTTTAAGCAAAATATATGTTGAGGATTTTGCAGTATCAATTTGTTAAAGATACTA 5940  
QY 5941 TACGTTGGCAAGTATTTTATTTATAGCTTACTAAATGTTTAAATACGGAACCAAGTCTATT 6000  
Db 5941 TACGTTGGCAAGTATTTTATTTATAGCTTACTAAATGTTTAAATACGGAACCAAGTCTATT 6000  
QY 6001 TTGCAAAATTTTAAATTTTAAAGAAATCTTTTAAAGAAATATTTTAACTTTGTTTAAAG 6060  
Db 6001 TTGCAAAATTTTAAATTTTAAAGAAATCTTTTAAAGAAATATTTTAACTTTGTTTAAAG 6060  
QY 6061 TATCTAAGCAAAATTTCTTTGCTTAAAGATTTTGTATAGAAATTTGTTTCGAAACCAAGTTT 6120  
Db 6061 TATCTAAGCAAAATTTCTTTGCTTAAAGATTTTGTATAGAAATTTGTTTCGAAACCAAGTTT 6120  
QY 6121 TTAAGAAAAATATATGTTTATAAGAGATATCATGATCTATTTAGTAAAAATTTCTTA 6180  
Db 6121 TTAAGAAAAATATATGTTTATAAGAGATATCATGATCTATTTAGTAAAAATTTCTTA 6180  
QY 6181 TAAATGTACCTATATATAATGTAGAAAAATTTTATCTAAATGTATAGATGATCAATTGTAA 6240  
Db 6181 TAAATGTACCTATATATAATGTAGAAAAATTTTATCTAAATGTATAGATGATCAATTGTAA 6240  
QY 6241 ATCAGACCTACAAACATATAGAGATTTCTTCTGGTGAATGACGGTAGTACGGTAATTCGG 6300  
Db 6241 ATCAGACCTACAAACATATAGAGATTTCTTCTGGTGAATGACGGTAGTACGGTAATTCGG 6300  
QY 6301 AAGAAATTTGTTTGTAGCATATCGAAGAAAGATAGTCGCATTTGTTTAAAAAAGAGA 6360  
Db 6301 AAGAAATTTGTTTGTAGCATATCGAAGAAAGATAGTCGCATTTGTTTAAAAAAGAGA 6360  
QY 6361 ACGGCGGCTATCAGATGCCCGTAATTTATGTCATTAAGTTCGGCCCAAGGTTGCTACTTTAG 6420  
Db 6361 ACGGCGGCTATCAGATGCCCGTAATTTATGTCATTAAGTTCGGCCCAAGGTTGCTACTTTAG 6420  
QY 6421 CTTTATAGATCAGATGATTTTATTTCAATTCGGAGTTTATCCAACTTTTACACGAGCA 6480  
Db 6421 CTTTATAGATCAGATGATTTTATTTCAATTCGGAGTTTATCCAACTTTTACACGAGCA 6480  
QY 6481 TTGAGAGAGAGATGCCCTTGTGGCAGTTGCTGTTATAGTAGGGTAGATGTTTGGGGC 6540  
Db 6481 TTGAGAGAGAGATGCCCTTGTGGCAGTTGCTGTTATAGTAGGGTAGATGTTTGGGGC 6540  
QY 6541 ATTTCTTAACGACGAGCGCTTCTTCAAAATCAGGCTGTTCTGAGCGGCGAGGAATGTTT 6600  
Db 6541 ATTTCTTAACGACGAGCGCTTCTTCAAAATCAGGCTGTTCTGAGCGGCGAGGAATGTTT 6600

```
QY 6601 GTAAAAAGCTGCTAGAGCGGATGCTCATCGCTTTGTGGTCCCTGTGTAATAAATCTCTATA 6660
Db 6601 GTAAAAAGCTGCTAGAGCGGATGCTCATCGCTTTGTGGTCCCTGTGTAATAAATCTCTATA 6660
QY 6661 AAAAAGAACTATTGGAAGATTTCGATTTGAAAAGGGTAAGATTTCATCAAGATGAATACT 6720
Db 6661 AAAAAGAACTATTGGAAGATTTCGATTTGAAAAGGGTAAGATTTCATCAAGATGAATACT 6720
QY 6721 TCATTATCGCTGCTCTATGAGTTAGAAAAAGTTGCAATAGTTAAGGAGTCTTTGTACT 6780
Db 6721 TCATTATCGCTGCTCTATGAGTTAGAAAAAGTTGCAATAGTTAAGGAGTCTTTGTACT 6780
QY 6781 ATTATGTTGACCGAGAAAATAGTATCAACAATCTTAGCATGACTGACCATCGTTCATTT 6840
Db 6781 ATTATGTTGACCGAGAAAATAGTATCAACAATCTTAGCATGACTGACCATCGTTCATTT 6840
QY 6841 GCCTACTGGAATTTCAAATGAACGAATGGACTTCTATGAAAGTAGAGAGATAAGAGC 6900
Db 6841 GCCTACTGGAATTTCAAATGAACGAATGGACTTCTATGAAAGTAGAGAGATAAGAGC 6900
QY 6901 TCTTACTAGAGTGTATCGTTCAATTTTAGCCTTTGCTGTTTGTGTTTAGGCAAATATA 6960
Db 6901 TCTTACTAGAGTGTATCGTTCAATTTTAGCCTTTGCTGTTTGTGTTTAGGCAAATATA 6960
QY 6961 ATCATTGCTGAGCAAAACAGCAAAAGAGCTT 6992
Db 6961 ATCATTGCTGAGCAAAACAGCAAAAGAGCTT 6992
```

## RESULT 2

```
US-09-767-041-29
; Sequence 29, Application US/09767041
; Patent No. US2002055168A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Hilda
; TITLE OF INVENTION: STREPTOCOCCUS SUIIS VACCINES AND DIAGNOSTIC TESTS
; FILE REFERENCE: 2183-4726
; CURRENT APPLICATION NUMBER: US/09/767,041
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: PCT/NL99/00460
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: EP98202465.5
; PRIOR FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: EP98202467.1
; PRIOR FILING DATE: 1998-07-22
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 29
; LENGTH: 6992
; TYPE: DNA
; ORGANISM: Streptococcus suis
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: CPS1
US-09-767-041-29
```

```
Query Match 100.0%; Score 6992; DB 9; Length 6992;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6992; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCGCCAAACGAAATGGCAATTTATTTGATATGATAGCAGTGGCAATTTCTGCAATCTTAA 60
Db 1 ATCGCCAAACGAAATGGCAATTTATTTGATATGATAGCAGTGGCAATTTCTGCAATCTTAA 60
QY 61 CAAGTCATATACCAAATGCTGATTTAAATCGTTCTGGAATTTTATCATATATGATGGTTTC 120
Db 61 CAAGTCATATACCAAATGCTGATTTAAATCGTTCTGGAATTTTATCATATATGATGGTTTC 120
QY 121 ATTTATTTGCAATTTTATATCTCGTATGCCAGTTGAAATTTGAGTATAGAGTAAATCTGA 180
Db 121 ATTTATTTGCAATTTTATATCTCGTATGCCAGTTGAAATTTGAGTATAGAGTAAATCTGA 180
QY 181 TAGAGTTTGA AAAACATTTAACTATAGTATAATATTGCAATTTTCTTACGGCAGTAT 240
```

```
Db 181 TAGAGTTTGA AAAACATTTAACTATAGTATAATATTGCAATTTTCTTACGGCAGTAT 240
QY 241 CATTTTGTTCGAGAAATAATTTCCGACATTTCAAGACGTTGGTCCCGTGTATTTCACATTAA 300
Db 241 CATTTTGTTCGAGAAATAATTTCCGACATTTCAAGACGTTGGTCCCGTGTATTTCACATTAA 300
QY 301 TAAACTTCGTTTTCGTTATACCTATTAAACGTAATTTATTAAGCAGTTTAAAGATAGCTTTC 360
Db 301 TAAACTTCGTTTTCGTTATACCTATTAAACGTAATTTATTAAGCAGTTTAAAGATAGCTTTC 360
QY 361 TATTTTCGACAACTATCAAAAAGAGCGATTTCTAATTTACAAACGCTGAACGATGGGAAA 420
Db 361 TATTTTCGACAACTATCAAAAAGAGCGATTTCTAATTTACAAACGCTGAACGATGGGAAA 420
QY 421 ATATGCAAGTTTATTTGAATCACATAAAACAAATTTCAAAAAAATCTTGTTCATTGGTAG 480
Db 421 ATATGCAAGTTTATTTGAATCACATAAAACAAATTTCAAAAAAATCTTGTTCATTGGTAG 480
QY 481 TTTTAGGTACAGAAATAGATAAAATTAATTTATCAATTAACCGCTCTATTATTTCTGTGGAAG 540
Db 481 TTTTAGGTACAGAAATAGATAAAATTAATTTATCAATTAACCGCTCTATTATTTCTGTGGAAG 540
QY 541 AAGCTATAGAGTTTCAACAGGGAAGTGGTCGACCACGCTCTTTATAAATCTTACCAAGTG 600
Db 541 AAGCTATAGAGTTTCAACAGGGAAGTGGTCGACCACGCTCTTTATAAATCTTACCAAGTG 600
QY 601 AGTTTTTAGAGCTAAAGCAATTCGTTTTCAGATTTTTCAGTTCGTAGTATTGATGTAAGCG 660
Db 601 AGTTTTTAGAGCTAAAGCAATTCGTTTTCAGATTTTTCAGTTCGTAGTATTGATGTAAGCG 660
QY 661 TTGATATTAAATTCATTCGTTTTCAGTTCGTTTGA AAAACAAAAATCAACTCTAGTGTG 720
Db 661 TTGATATTAAATTCATTCGTTTTCAGTTCGTTTGA AAAACAAAAATCAACTCTAGTGTG 720
QY 721 ACCATAGCATTTGTAACCTTTTCCAAATTTTATAAGCCTAGTTCATATCATGATGAAC 780
Db 721 ACCATAGCATTTGTAACCTTTTCCAAATTTTATAAGCCTAGTTCATATCATGATGAAC 780
QY 781 GACTTTTGGATATATCTCGAGCGGTAGTCGGTGTAAATTTTGTGGTATAGTTCTATT 840
Db 781 GACTTTTGGATATATCTCGAGCGGTAGTCGGTGTAAATTTTGTGGTATAGTTCTATT 840
QY 841 TGTAGTTTCCAAATTTTCGTAGAGATGTTGACCGGCTATTTTGTCTCAGAAACGAGTTG 900
Db 841 TGTAGTTTCCAAATTTTCGTAGAGATGTTGACCGGCTATTTTGTCTCAGAAACGAGTTG 900
QY 901 GACAGATGGACGCAATTTTACATTTCAAGTTTTCGATTCGATGATGTTGATGCTGAGG 960
Db 901 GACAGATGGACGCAATTTTACATTTCAAGTTTTCGATTCGATGATGTTGATGCTGAGG 960
QY 961 AGCGCAAAAAGACTTCTCGACCAAAACAGATGCAAGGGTGGGTATGTTTAAATCG 1020
Db 961 AGCGCAAAAAGACTTCTCGACCAAAACAGATGCAAGGGTGGGTATGTTTAAATCG 1020
QY 1021 GAAAAACGATCCTAGAAATTTACTCCAAATTTGGACATTTTCATAGCAAAAAACAGTTTAGACG 1080
Db 1021 GAAAAACGATCCTAGAAATTTACTCCAAATTTGGACATTTTCATAGCAAAAAACAGTTTAGACG 1080
QY 1081 AGTTACCAAGTTTATTAATGTTTAAATTTGCGGATATGAGTCTAGTTGGTACACTCCAC 1140
Db 1081 AGTTACCAAGTTTATTAATGTTTAAATTTGCGGATATGAGTCTAGTTGGTACACTCCAC 1140
QY 1141 CTACAGTTGATGAATTTGAAAAATATATCTCTGGTCAAAAGAGAGATGAGTTTAAAC 1200
Db 1141 CTACAGTTGATGAATTTGAAAAATATATCTCTGGTCAAAAGAGAGATGAGTTTAAAC 1200
QY 1201 CAGGATTTACAGGCTCTCTGGCAGGTTAGTGGTCTAGTAAATATACAGACTTCGACGAGC 1260
Db 1201 CAGGATTTACAGGCTCTCTGGCAGGTTAGTGGTCTAGTAAATATACAGACTTCGACGAGC 1260
QY 1261 TAGTTCGGTTCGACTTAGCATACATTTGATAATTTGGACTATCTGGTCAAGATTTTAAATTT 1320
Db 1261 TAGTTCGGTTCGACTTAGCATACATTTGATAATTTGGACTATCTGGTCAAGATTTTAAATTT 1320
```

Db 1261 TAGTTGGTTCGACCTTAGCATACATTCGATTAATTCGACTATCTCGTCAGATATTAATAATTT 1320  
Qy 1321 TATTAAAGACAGTGAAGTTGTATTTGTGTGAGAGAGGGAAGTAAGTAAAGATATATGAAAG 1380  
Db 1321 TATTAAAGACAGTGAAGTTGTATTTGTGTGAGAGAGGGAAGTAAGTAAAGATATATGAAAG 1380  
Qy 1381 TTTTGTGTTGGTTCGTTCTTTCAGGGGACATTTGACTCACTCTGTTATTTCTTAAACCGTTTT 1440  
Db 1381 TTTTGTGTTGGTTCGTTCTTTCAGGGGACATTTGACTCACTCTGTTATTTCTTAAACCGTTTT 1440  
Qy 1441 GGAAGGAAGAAACGTTTTTGGGTAAACATTTGATAAGAGGATGCAAGAAAGTCCTTTGA 1500  
Db 1441 GGAAGGAAGAAACGTTTTTGGGTAAACATTTGATAAGAGGATGCAAGAAAGTCCTTTGA 1500  
Qy 1501 AGAATGAAAAATGTATCCATGTTACTTTCACAAACAAATCGCAATCTCATTTAAATTTAGTGA 1560  
Db 1501 AGAATGAAAAATGTATCCATGTTACTTTCACAAACAAATCGCAATCTCATTTAAATTTAGTGA 1560  
Qy 1561 AAAATACCTTTCTAGCTTTCAAAATTTTACGTGATGAGAAACCAAGATGTTATTTTCAT 1620  
Db 1561 AAAATACCTTTCTAGCTTTCAAAATTTTACGTGATGAGAAACCAAGATGTTATTTTCAT 1620  
Qy 1621 CTGTTGGCGGCTGCTGCTCCCTTCTTTTACATCGGAAACTATTTGGAGCAAGACGA 1680  
Db 1621 CTGTTGGCGGCTGCTGCTCCCTTCTTTTACATCGGAAACTATTTGGAGCAAGACGA 1680  
Qy 1681 TTTATATGAAAGTATTTGATCGAGTTAATAAATCTACATTAACCTGGAACCACTAGTTTATC 1740  
Db 1681 TTTATATGAAAGTATTTGATCGAGTTAATAAATCTACATTAACCTGGAACCACTAGTTTATC 1740  
Qy 1741 CCGTAAACAGATATTTTATTTGTTTCAGTGGGAGAAATGAAAGAGGTATATCTTAAATCTA 1800  
Db 1741 CCGTAAACAGATATTTTATTTGTTTCAGTGGGAGAAATGAAAGAGGTATATCTTAAATCTA 1800  
Qy 1801 TTAATCTGGGAGTATTTTAAATGATTTTGTAAACAGTAGGAACTCATGAACAAACAGTT 1860  
Db 1801 TTAATCTGGGAGTATTTTAAATGATTTTGTAAACAGTAGGAACTCATGAACAAACAGTT 1860  
Qy 1861 TAATCGATTGATAAAGAGATTTGATTTTATTTGAAAAAAAATGGAAGTATAACCGACGAAAT 1920  
Db 1861 TAATCGATTGATAAAGAGATTTGATTTTATTTGAAAAAAAATGGAAGTATAACCGACGAAAT 1920  
Qy 1921 ATTTTATTCACACAGATATCTGACTATATTCAGATATTCGAGATATGCAAGTATAAAAAATTTCT 1980  
Db 1921 ATTTTATTCACACAGATATCTGACTATATTCAGATATTCGAGATATGCAAGTATAAAAAATTTCT 1980  
Qy 1981 CAGTTACAAAGAAATGGAACAAATATATTAACAAATCAGAAGTAGTATTTGCCACGGAGG 2040  
Db 1981 CAGTTACAAAGAAATGGAACAAATATATTAACAAATCAGAAGTAGTATTTGCCACGGAGG 2040  
Qy 2041 CCCCCTACTTTTATGAATTCATTTATCAAGGAAAAAACAATATTTGTTTCTCTAGACA 2100  
Db 2041 CCCCCTACTTTTATGAATTCATTTATCAAGGAAAAAACAATATTTGTTTCTCTAGACA 2100  
Qy 2101 AAAAAAGTATGCTGAACATGTAATGATCATCAAGTAGAGTTTGTAGAGAAATTTTACA 2160  
Db 2101 AAAAAAGTATGCTGAACATGTAATGATCATCAAGTAGAGTTTGTAGAGAAATTTTACA 2160  
Qy 2161 AGATAAATAATTTTATTTATAGAAAAATATAGATGATTTGTTGAAAAAATTTATGAAAT 2220  
Db 2161 AGATAAATAATTTTATTTATAGAAAAATATAGATGATTTGTTGAAAAAATTTATGAAAT 2220  
Qy 2221 TTCTAAGCAAACTAACTTACATCAAAATAATAATTTTGTGAAAAAGATTAATAACAAAT 2280  
Db 2221 TTCTAAGCAAACTAACTTACATCAAAATAATAATTTTGTGAAAAAGATTAATAACAAAT 2280  
Qy 2281 AGTTGAAAAATTTTATGAGGATCAAGAAAAATGAATAATAAAAAAGATGCGATTTTGATAA 2340  
Db 2281 AGTTGAAAAATTTTATGAGGATCAAGAAAAATGAATAATAAAAAAGATGCGATTTTGATAA 2340  
Qy 2341 TGGCTTATCAATAATTTTCTCAGATTTTACCTGGAGAGGATACAGATATTTATCATCTTCT 2400  
Db 2341 TGGCTTATCAATAATTTTCTCAGATTTTACTGGAGAGGATACAGATATTTATCATCTTCT 2400

Qy 2401 CTCAGAGAAATGCACACCAATTAGTTCTCTCAGAAATACCTGTATATAATTTATTTAAATATTC 2460  
Db 2401 CTCAGAGAAATGCACACCAATTAGTTCTCTCAGAAATACCTGTATATAATTTATTTAAATATTC 2460  
Qy 2461 TCAGGAATTTATATGTTGAAATTTACAAAAAGATGAGCAAAAAATATAAAGAAAAATAGGATATA 2520  
Db 2461 TCAGGAATTTATATGTTGAAATTTACAAAAAGATGAGCAAAAAATATAAAGAAAAATAGGATATA 2520  
Qy 2521 TGAACGAGTTAAATGTTTACAGATTTATTTCTCTAATATATTCAGAAAAAATCTATTGATAATGT 2580  
Db 2521 TGAACGAGTTAAATGTTTACAGATTTATTTCTCTAATATATTCAGAAAAAATCTATTGATAATGT 2580  
Qy 2581 ACTGTTTAGAATTTTATTAAGAATGTATCGAGCTTTTGAATCTACTATTTTACAAGATTTGTT 2640  
Db 2581 ACTGTTTAGAATTTTATTAAGAATGTATCGAGCTTTTGAATCTACTATTTTACAAGATTTGTT 2640  
Qy 2641 GTTTATTTGATAGAAATAAAAAATGTTCTAAGAAATGAAGTTTGGTCTTAATTTGGGTTTCG 2700  
Db 2641 GTTTATTTGATAGAAATAAAAAATGTTCTAAGAAATGAAGTTTGGTCTTAATTTGGGTTTCG 2700  
Qy 2701 CTTTCCACATGATTTTGTGGCAATTTCTTTTATCAAAATGAAACGAAACAGCTTATTTATTT 2760  
Db 2701 CTTTCCACATGATTTTGTGGCAATTTCTTTTATCAAAATGAAACGAAACAGCTTATTTATTT 2760  
Qy 2761 AAGTAATCTAAATGTCGAGATGAACATTTTATACAGACAATTTATAGAAAAATATGAATTT 2820  
Db 2761 AAGTAATCTAAATGTCGAGATGAACATTTTATACAGACAATTTATAGAAAAATATGAATTT 2820  
Qy 2821 TCAAAATAGATTTATCTAAATATGAAATTTAAGATATATAAAGTGGAAAAAATCAACATCT 2880  
Db 2821 TCAAAATAGATTTATCTAAATATGAAATTTAAGATATATAAAGTGGAAAAAATCAACATCT 2880  
Qy 2881 TCTCTTATTTGCTTTACAGATGATTTCTATTGATGAATTTGCTAAATGCAAGAAATTTTAGGT 2940  
Db 2881 TCTCTTATTTGCTTTACAGATGATTTCTATTGATGAATTTGCTAAATGCAAGAAATTTTAGGT 2940  
Qy 2941 TTTTATTTGCTAGAAAGTTAAAAATAGAAAAATGAAAAATGAAAAATTTTAAAGAAAAATTTACT 3000  
Db 2941 TTTTATTTGCTAGAAAGTTAAAAATAGAAAAATGAAAAATGAAAAATTTTAAAGAAAAATTTACT 3000  
Qy 3001 AAAAAATAAAAATGATTTTCTGAGAGTAAGTATGTTTAAATTTAAATTTAAAAATATGACC 3060  
Db 3001 AAAAAATAAAAATGATTTTCTGAGAGTAAGTATGTTTAAATTTAAATTTAAAAATATGACC 3060  
Qy 3061 CGGAATATTTTATTTTAAAGTACTTCTGTTGATTTATTTTATTTTATTTTCCAGAGCAAAAGTATG 3120  
Db 3061 CGGAATATTTTATTTTAAAGTACTTCTGTTGATTTATTTTATTTTATTTTCCAGAGCAAAAGTATG 3120  
Qy 3121 TATTTTATTTAAATTTTATGAAATTTAAATTTTATTTTATTTTAAATTTTAAAAATCTAAGC 3180  
Db 3121 TATTTTATTTAAATTTTATGAAATTTAAATTTTATTTTATTTTAAATTTTAAAAATCTAAGC 3180  
Qy 3181 TAATATTTAAAAATGAAATTTTATGTTTATTTATGTTTATTTATTTTATTTTATTTTCCAG 3240  
Db 3181 TAATATTTAAAAATGAAATTTTATGTTTATTTATGTTTATTTATTTTATTTTATTTTCCAG 3240  
Qy 3241 TAGTCACAGTATGTTTGTGAAATTTAAATTTTGAAGATTTATTTGCAAGATTTTACTGCTC 3300  
Db 3241 TAGTCACAGTATGTTTGTGAAATTTAAATTTTGAAGATTTATTTGCAAGATTTTACTGCTC 3300  
Qy 3301 CCATAATTTGGATTTTGCATTAATGTAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAAT 3360  
Db 3301 CCATAATTTGGATTTTGCATTAATGTAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAAT 3360  
Qy 3361 ATAAAAATTTAAAAATAGTATCTTTTATGTTTATTTAGTTTATTTAGGATATCTGAT 3420  
Db 3361 ATAAAAATTTAAAAATAGTATCTTTTATGTTTATTTAGTTTATTTAGGATATCTGAT 3420  
Qy 3421 TGTATATTTTCAAAATGGGAAAGATATTTGATTTTTTTTAGACAGACACCTTTATAGGACTAG 3480  
Db 3421 TGTATATTTTCAAAATGGGAAAGATATTTGATTTTTTTTAGACAGACACCTTTATAGGACTAG 3480







```
US-10-865-873-3
; Sequence 3, Application US/10865873
; Publication No. US20050064559A1
; GENERAL INFORMATION:
; APPLICANT: Miyake, Katsumide
; APPLICANT: Watanabe, Masaki
; APPLICANT: Iijima, Shinji
; TITLE OF INVENTION: Beta 1,3-galactosyltransferase and DNA encoding the same
; FILE REFERENCE: 786.53
; CURRENT APPLICATION NUMBER: US/10/865,873
; CURRENT FILING DATE: 2004-06-14
; PRIOR APPLICATION NUMBER: US/09/900,038
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: JP 2001-392
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 6865
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae type Ib
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (617)..(1789)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1816)..(2262)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2265)..(2744)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2843)..(3979)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3982)..(4953)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (5009)..(5947)
; US-10-865-873-3

Query Match 10.4%; Score 727.8; DB 19; Length 6865;
Best Local Similarity 59.2%; Pred. No. 9.1e-93;
Matches 1389; Conservative 0; Mismatches 917; Indels 41; Gaps 7;

Qy 16 TGGCATTATTTGATGATAGCAGTTGCAATTTCTGCAATCTTTAAACAAGTCATATACCAA 75
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
438 TAGCGATGATTCAAACCGTTGTGGTTTATTTCTGCAAGTTTGACATTAACATTAATTA 497
Qy 76 ATGCTGATTTAAATCG---TTCTGGAATTTTATCATATGATGATGTTTCATTTTTCAT 132
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
498 CTCCCAACTTTAAAGCAATAAAGATTTATTTGTTGTTCTATTGATACATTTATTTGCT 557
Qy 133 TTTTATATCTCGATGCGAGTTCAATTTTGATATAGAGGTAATCTGATAGAGTTTCGAAA 192
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
558 TTTATCTTTCGATTTTACAGAGCTTTTGAGTGTGGCTATCTTGAAGAGTTTAAAA 617
Qy 193 AAACATTAACATATAGTATATATTTGCAATTTTCTTACGGCAGTATCATTTTGTGTG 252
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
618 TGGTATTGAATACAGCTTTTACTATATTTTTCATATCAAGTTTCATTTTATTTTCTA 677
Qy 253 AGAATAATTTCCGACTTTCAAGAGCGTGGCCGTTATTTTACATTAATTAACCTTCGTTT 312
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
678 AAAAATCTCTTTTACACGACAGCTTTCTTTTTTACTTTTATTTGCTATGAAATTCGATTT 737
Qy 313 TGGTATACCTATTAACTTAATTAAGCAGTTTAAAGGATAGCTTTCTATTTTTCGACAA 372
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
738 TATTATATCTATGAAATTCATTTTAAATATATATTCGAAATATCTTACGCTAAGTTT 797
Qy 373 TCTATCAAAAAAGACCGATTTCTAATTAACAACGGCTGAACGATGGGAAAAATATCAAGTTT 432
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
798 CACGAGATACCAAGTTGTTTGTGATAACGAATAAGGATTTCTTTTATCAAAATGACCTTTA 857
Qy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

433 TATTTGAATCACATAAACAATTCAAAAAATCTTCTGTCATTGATGTTTATAGGTACAG 492
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
858 GGAATAAAT---ACGACCATAATATATCGCTGCTGTATCTTGGATTCCTCTGAAAAGG 914
Qy 493 AAATAGATAAAAAATTAATTTATCATTTACCGCTCTATTATTCTGTGGGAAGACTATAGAGT 552
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
915 ATTGTTATGATTTGAACAATACTCGTTAAGGATAATAACAAGATGCTCTTACTTTCAG 974
Qy 553 TTTCAACAAGGGAGTGGTCGACCAAGCTCTTTATAAATCTACCAAGTGAAGTTTATAGACG 612
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
975 AGTTAACTCTGCTTAACTGTTGATCAAGCTTTTATTAACATACCAATTAATTAATTTGGTA 1034
Qy 613 TAAAGCAATTC-----GTTTCAGATTTTCAGTTGTTAGGTATTGATGAAGCGTTG 663
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1035 AATACCAAAATACAAGATATTTAATGACATTTGAAGCAATGGGAGTGAATGTCAATGTTA 1094
Qy 664 ATATTAAATTCATTCGGTTTACTGCTGTGAAAAAACAATAAATCCAACTCTAGGTGACC 723
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1095 ATGTAGAGGCACTTAGCTTTGATAATATAGGAGAAAGCGAATCCAACTTTTGAAGGAT 1154
Qy 724 ATAGCATTTGTAATTTTTCACAAAATTTTATAAGCCTAGTCTATATCATGATGAACAGAC 783
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1155 ATAGTGTATTACATATTTCTATGAAATTTCTATAAATATAGTCACTTATAGCAAAACGAT 1214
Qy 784 TTTTGGATATACTCGAGCGGTAGTCGGGTTAATTTTGTGTATAGTTTCTATTTTGT 843
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1215 TTTTGGATATCATGGGTGCTATTATAGGTTTGTCTATATGTGGCATTTGGCAATTTTTC 1274
Qy 844 TAGTTCCAATTTATTCGTAGAGATGGTGGACCGGCTATTTTGTCTCAGAAACGATTTGGAC 903
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1275 TAGTTCCGCAATCAGAAAAGATGGTGGACCGGCTATCTTTCTCAAAATAGATAGTGT 1334
Qy 904 AGAATGAGCGCATTTTACATTTCAAGTTTTCGATGATGATGATGATGATGATGATGATG 963
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1335 GTAAATGTTAGGATTTTGTAGATTCTATAAATTCAGATCAATGCGAGTAGATGCGAACA 1394
Qy 964 GCAAAAAAGACTTGTCTAGCCAAAACAGATGCAAGGTGGGTATGTTTAAATGGGAA 1023
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1395 TTAAGAAAGATTTTATTTAGTTTCAACATCA-----AATGACGGGCTAATGTTAAGTTAGA 1449
Qy 1024 AAACGATCTTAGAATTTACTCCAATTTGACATTTTATACGCAAAAAACAAGTTTATAGACG 1083
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1450 CGATGATCTTAGAATTTCTAAATATAGAAAATTTATTCG-AAAAACNAGCATAGATGAGT 1508
Qy 1084 TACCAAGCTTTTATAATGTTTAAATTTGGGATATGATGCTAGTTGGTACAGTCCACCTA 1143
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1509 TGCCTCAATTTCTATAATGTTTAAAGGTGATATGATGTTTGTAGTAAACAACGCGCTCCCA 1568
Qy 1144 CAGTTGATGAATTTGAAAAATATCTCTGCTCAAAAGAGACGATTTGAGTTTAAACACG 1203
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1569 CAGTTGATGAATATGAAAAGTATAATTTCAACGCAAGAGCGACGCTTAGTTTAAAGCCAG 1628
Qy 1204 GGAATACAGCTCTCTGGCAGGTTAGTGGTCTAGTAAATATACAGACTTTCGACACGTAG 1263
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1629 GAATCACTGGTTGTGCAAAATATCTGTTAGAAATATATTAATCTGATTTTGTGAAATCG 1688
Qy 1264 TCCGTTGGACTTTAGCATACATTTGATAATTTGGAATATCTGCTGATATTTAAATTTTAT 1323
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1689 TAAAGTTAGATGTTTCAATATATCAATGAATGGTCTATTTGTTGTCAGATATTAAGATTATTC 1748
Qy 1324 TAAAGCAGTGAAGTTTGTATTTGTTGAGAGAGGAGTAAAGTAAAGTAAAGTAAAGTAAAGT 1372
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1749 TCCTAACGCTTAAAGGTTTACTCGGGAACAGGAGCTAAGTAAAGTAAAGTAAAGTAAAGTAAAG 1808
Qy 1373 -----TATGAAAGTTTGTGTTGCTCGGTTCTTCCAGGGGACATTTGACTCACTTTGATTTT 1426
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1809 GAATATAATGAAATTTTGTCTGTTGTTCAAGTGGTGGTCCCTAGCACACACTTGAACCT 1868
Qy 1427 GTTAAACCGTTTTTGAAGAGAGAAACGTTTTTGGGTTAAACATTTTGAATGAAGAGATGC 1486
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1869 TTTGAAACCCATTTGGGAAAAAAGAGATAGTTTTTGGGTAACTTTTGTGATGAAGAGATGC 1928
Qy 1487 AAGAAGTCTTTTGAAGAAATGAAAAAATGTATCCATGTTACTTTTCCCAAAATCGCAATCT 1546
```



```
Db 1929 TAGGAGTATCTTAAGAGAGAGATGTTATATCATTTGCTTCTTCCAAACAAACCGTAATGT 1988
Qy 1547 CATTAAATTTAGTGAATAATCTTTCTTAGCTTTTCAAAATTTTACGCTGATGAGAACCCAGA 1606
Db 1989 CAAAACCTTGGTAAATAATCTATTTCTAGCTTTTAAAGGTCCTTAGAAAGAAGAACCCAGA 2048
Qy 1607 TGTTATTTATTCATCTGCTGGCGGCTTGTCTGCTCCCTCTTTTACATCGGAAACTATTT 1666
Db 2049 TGTTATCATATCATCTGCTGGCGGCTGTAGCAGTACCATTTCTTTTATATTTGGTAAGTTAT 2108
Qy 1667 TGGAGCAAGACGATTTATTTATTTGAAGTATTTTGATCGAGTTAATAATCTACATTAACCTGG 1726
Db 2109 TGGCTGTAAAGACCGTTTATATATAGAGTTTTCGACAGGATAGATAACCACTTTGACAGG 2168
Qy 1727 AAAAAGTATTTATCCCGTAAACAGATATTTTATTTGTTCCAGTGGGAAGAAATGAAGAAGGT 1786
Db 2169 AAAATAGTGTATCTCTGAACAGATAAATTTATTTGTTCCAGTGGGAAGAAATGAAGAAAGT 2228
Qy 1787 ATATCTTAATCTATTAATCTTGGGAGATTTTATTTTAAATGATTTTGTAAACAGTAGGAAGT 1846
Db 2229 TTATCTTAAGCAATTAATTTAGGAGGAATTTTAAATGATTTTGTCCACAGTAGGAGCA 2288
Qy 1847 CATGAACACAGTTTAATCGATTTGATAAAGAGATTTTATTTCAAAAAAATGGAAGT 1906
Db 2289 CATGAACAGCAGTTCAACCGCTCTTATTAAGAAGTTGATAGATTAAAGGAGCAGGTGCT 2348
Qy 1907 ATAACCGACGAAATATTTATTTCAACAGGATTTCTGACTATATTTCCAGATATTTGCAAG 1966
Db 2349 ATTGATCAGAAGTGTTCAATTCACCGGTTACTCAGACTTTGAACCTCAGATTTGTGAG 2408
Qy 1967 TATAAAAAATTTCTCAGTTTACAAAGAAATGGAACAATATATTAACAAATCAGAAGTAGTT 2026
Db 2409 TGGTCAAAATTTCTCATATGATGATATGAATGAATCTTACATGAAAGAGCTGAGATTGTT 2468
Qy 2027 ATTTGCCACGAGCGCCGCTACTTTTATGATTTTATGATTTCAATTTCCAAAGGAAAAACAATTA 2086
Db 2469 ATCACACACGCGGCTCCAGACGCTTTATGATGAGTGAATGAGTCTTCAAGGGAAGAAACTAT 2528
Qy 2087 TTGTTTCTTAGCAAAAAAAGTATGTCGCAACATGTAATGATCATCAAGTAGAGTTTCTGTA 2146
Db 2529 GTGGTCTTAGCAAGAACAGTTTGGAGAGCATGTGAATATCATCAGTGAGATTTTGTG 2588
Qy 2147 AGAAGAAATTTTACAGATAATAATATTTTATTTATAGAAAAATATAGATGATTTGTTGAA 2206
Db 2589 AAAGAGTTATTTCTTGAATATGAGTTAGATTATATTTTGAATATCAGTGAATTAGAGAT 2648
Qy 2207 AAAATTATGAGTTTCTAAGCAAC---TAACTTTACATCAATATAATTTTATTTTGT 2263
Db 2649 ATTATTAAGGAAAAAATATATCTACTAGTAAAGTAATATATCACAAAAAATGATTTTGT 2708
Qy 2264 GAAAGATTAAACAAATAGTTGAAAAATTTAATGAGGATCAAGAAAAATGAAATAATAAAA 2323
Db 2709 TCCTCTTCAAAATGAACCTTCTAAACTATTTTGAATAAATATATTTTGTGGAGAAAAA 2768
Qy 2324 AGATGCA 2330
Db 2769 AATTGAA 2775
```

RESULT 5  
US-09-870-759-83  
; Sequence 83, Application US/09870759  
; Patent No. US20020177551A1  
; GENERAL INFORMATION:  
; APPLICANT: TERMAN, David S  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE  
; FILE REFERENCE: 870759  
; CURRENT APPLICATION NUMBER: US/09/870,759  
; CURRENT FILING DATE: 2002-01-14  
; PRIOR APPLICATION NUMBER: US 60/208,128  
; PRIOR FILING DATE: 2000-05-30  
; NUMBER OF SEQ ID NOS: 166

; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 83  
; LENGTH: 17276  
; TYPE: DNA  
; ORGANISM: Streptococcus agalactiae  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (7062)..(8207)  
; OTHER INFORMATION:  
US-09-870-759-83

Query Match 10.4%; Score 726.8; DB 9; Length 17276;  
Best Local Similarity 59.2%; Pred. No. 1.7e-92;  
Matches 1366; Conservative 0; Mismatches 902; Indels 38; Gaps 6;

```
Qy 16 TGGCATTATTTGATATGATAGTAGGTTGCAATTTTCTGCAATCTTAAACAAGTCATATACCAA 75
Db 4765 TAGCGATGATTCANACAGTTCTAGTTTATTTTCTGCAAGTTTGACATTAACATTAATTA 4824
Qy 76 ATGCTGATTTAAATCG---TTCTGGAAATTTTATCATAAATGATGGTTCATTTATTTTGCAT 132
Db 4825 CTCCCAATTTTAAAGCAATAAAGATTTATTTGTTTGTCTATTTGATACATATATTTGTTT 4884
Qy 133 TTTTATATCTCGTATGCCAGTTTCAATTTGAGTATAGAGTTAACTGTATAGATTTTGA 192
Db 4885 TTTATCTTTCTGATTTTACAGAGACTTTTGGAGTCGTCGTATCTTGAAGAGTTTAA 4944
Qy 193 AAACTATTAACATAGTATATATTTTCAATTTTCTTACGGCAGTATCATTTTGTGG 252
Db 4945 TGGTATTTGAATACAGCTTTTACTATATTTTTCATATCAAGTTTCATTTTATTTTATTTTA 5004
Qy 253 AGAATAATTTTCGACTTTTCAAGACGTGGTGGCGGTATTTTCCACATTAATAAATTCGTTT 312
Db 5005 AAAACTTTTACACAGACAGCTTTTCTTTTACTTTATTTTATTTTCTATGAATTCGATTT 5064
Qy 313 TGGTATACCTATTTAAACGTAATTAAGCAGTTTAAAGGATAGCTTTCTATTTTTCGACA 372
Db 5065 TATTATATCTATTTGAATTTCAATTTTAAATAATTTATCGAAAAATATTTCTTACGTAAGTTT 5124
Qy 373 TCTATCAAAAAAGACGATTTCTAATTAACACGGCTGACGATGGGAAATATGCAAGTTT 432
Db 5125 CACGAGATACCAAGTTTGTGTTTGAATACGAATAAGGATTTCTTATCAAAATGACCTTTA 5184
Qy 433 TATTTGAATCACATAAAACAAATTTCAAAAAATCTTTTGTGCATTTGTTAGTATAGTACAG 492
Db 5185 GGAATAAT---ACGACCAATAATTTATATCGCTGCTGTATCTTCGACTCCTCTGAAAGG 5241
Qy 493 AAATAGATAAAATTAATTTATCATTTACCGCTCTATTTATTTCTGTGGAGAGCATAGAGT 552
Db 5242 ATTGTTATGATTTGAAAACATAACTCGTTAAGGATAATAAAACAAAGATGCTCTTACTTCAG 5301
Qy 553 TTTCAACAAGGAGTGGTCGACCACTGCTTTTATAAATCTACCAAGTCAGTTTGTAGACG 612
Db 5302 AGTTAACTGTTTAACTGTTGATCAAGCTTTTATTAACATACCCATTTGAAATTTTGGTA 5361
Qy 613 TAAAGCAA-----TTCTGTTTTCAGATTTTGTAGTTTGTAGGTTATTCATGTAAGCGTTG 663
Db 5362 AATACCAATACAGATATTTAATGACATTTGAAGCAATCGGAGTGTGTCATGTTA 5421
Qy 664 ATATTAATTCATTCGGTTTACTCGGTTGAAAAACAAAAATCCAACTGCTAGTGTACC 723
Db 5422 ATGTAGAGGCACCTTAGCTTTGATAATATAGGAGAAAGCAATCCAACTTTTGAAGGAT 5481
Qy 724 ATAGCATTTGAACCTTTTCCCAAAATTTTATAGCCTAGTTCATATCATGATGAACGAC 783
Db 5482 ATAGTGTATTACATATTTCTATGAATTTCTATNAATATATAGTCACCTTATAGCAAAACGAT 5541
Qy 784 TTTTGGATATACTCGGAGCGGTAGTCGGTTTAAATTTATTTTGGTATAGTTTCTTATTTGT 843
Db 5542 TTTTGGATATACCGGTGCTATTAATAGTTTGTCTCATATGTTGGCATTTGTGCAATTTTTC 5601
Qy 844 TAGTTCCAAATTTATTCGTAGAGATGGTGACCGGCTATTTTGTCTCAGAAACGAGTTGAC 903
```



Db 5602 TAGTTCGCAATCAGAAAGAGATGGTGGACGGCTATCTTTTCTCAAAATAGAGTAGGTC 5661  
Qy 904 AGAATGACGCAATTTACATCTTACAAAGTTTCGATCGATGTATGTTGATGCTGAGGAGC 963  
Db 5662 GTAATGCTAGGATTTTATGATTTCTATTAATTCAGATCAATGCGAGTAGATGCGAAACAAA 5721  
Qy 964 GCAAAAAGACTTGCTCAGCCCAAAACAGATGCAAGGGTGGGTATGTTTAAATGGGAA 1023  
Db 5722 TTAAGAAGATTTTATTAAGTTACATCAATCAATGACAGGG-----CTAATGTTAAGTTAGA 5776  
Qy 1024 AAACGATCTAGATTAATCTCAATTTGACATTTTCATACGCAAAACAAAGTTTATAGACAGT 1083  
Db 5777 AGATGATCCTAGATTAATCTCAATTTAGGAAATTTATTCG-AAAAACAAGCATAGATGAGT 5835  
Qy 1084 TACCACAGTTTATAATGTTTAAATGCGGATATGACTAGTGTGTTGATGACGCTCCACCTA 1143  
Db 5836 TGCCTCAATTTCTAATATGTTTAAAGGCGATATGAGTTTAGCAGGAACACGCGCTCCCA 5895  
Qy 1144 CAGTTGATGAATTTGAAAATATATCTCTGGTCAAAAGACGATTTGAGTTTAAACCCAG 1203  
Db 5896 CAGTTGATGAATATGAAAAGTATTAATTTCAACGCAAGACGACGCTTAGTTTAAAGCCAG 5955  
Qy 1204 GGATTACAGTCTCTGCGCAGTTAGTGTCTAGTATATATACAGACTTCGACGACGTAG 1263  
Db 5956 GAATCACTGTTTGTGCAAAATATCTGTAGAAATATATATCTGATTTTGTGAATCG 6015  
Qy 1264 TTCGGTTGGACTTAGCATAATTTGATTAATTTGACTCTCTGTCAGATTAATTTAAATTTAT 1323  
Db 6016 TAAAGTTAGATGTTTCAATATATCAATGAATGGTCTATTTGGTCAGATTAATTAAGATTATTC 6075  
Qy 1324 TAAAGACAGTGAAGTTGTATTTGTTGAGAGGGAAGTAAAGTA----- 1372  
Db 6076 TCCTAAACACTAAAGGTAGTTTACTCGGGAACAGAGCTAAGTAAAGTAAAGTTTGAAG 6135  
Qy 1373 -----TATGAAGTTGTTTGGTCGGTCTTTCAGGGGACATTTGACTCACTTGTATTT 1426  
Db 6136 GAATATATGAAATTTGTCGGTTGGTTCAAGTGGTGGTCACTAGCACACTTGAACCT 6195  
Qy 1427 GTTAAACCGTTTGGGAAGGAAGAACGTTTTTGGGTAAACATTTTGTATAAGAGGATGC 1486  
Db 6196 TTTGAAACCCATTTGGGAAAGGAAGATAGGTTTTTGGGTAAACCTTTTATAAAGAGATGC 6255  
Qy 1487 AAGAGTCTTTTGAAGATGAAAATGTATCCATTTACTTCCCAACAAATCGCATCT 1546  
Db 6256 TAGGATTAATCTAAGAGAGAGATTTGATATCAATTTGCTTTCTTCCAAACAAACCGTAAATGT 6315  
Qy 1547 CAATTAATTTAGTGAATAATCTTTCTTAGCTTTTCAAAATTTTACGTGATGAGAAACGAGA 1606  
Db 6316 CAAAACCTTGGTAAATAATTAATTTCTAGCTTTTAAAGTCTTTAGAAAAGAGAACCGAGA 6375  
Qy 1607 TGTATTATTTCATCTCGTGGCGGCTGCTGTCCTTCTTTTACATCGGAAACATTT 1666  
Db 6376 TGTATCATATCATCTGCTGCGCTGTAGCAGTACCATTCTTTTATATTTGTTAAGTTATT 6435  
Qy 1667 TGGAGCAAGACGATTTATTTAGAGTATTTTATGATGAGTAAATTAATCTACATTAACG 1726  
Db 6436 TGGTTGTAAGACCGTTTATATAGAGGTTTTCGACAGGATAGATAAACCACTTTTGAAG 6495  
Qy 1727 AAAACTAGTTTATCCCGTAAACAGATATTTTATTTGTTTCAAGTGGGAAGAAATGAAGAGT 1786  
Db 6496 AAAATTAGTATCTCTGTAACAGATAAATTTATTTGTTTCAAGTGGGAAGAAATGAAGAAAGT 6555  
Qy 1787 ATATCTTAATCTATTAACTTTGGGAGTATTTTTTAAATGATTTTTTGTAAACAGTAGGAAC 1846  
Db 6556 TTATCTTAAGCAATTAATTTAGGAGGAAATTTTAAATGATTTTTTGTACAGTGGGACA 6615  
Qy 1847 CATGAACACAGTTTAAATCGATTGATAAAGAGATTTGATTTTGTAAAGAAATTTGAAGT 1906  
Db 6616 CATGAACAGGAGTTTCAACCGCTTTATTTAAAGAGTTTATAGATTTAAAGGGACAGGTGCT 6675  
Qy 1907 ATAAACGACGAAATATTTTATTTCAACAGGATATTTCTGACTATATTTCCAGAAATTTGCAAG 1966  
Db 6676 ATTGATCAAGAGTGTTCATTTCAACCGGTTTACTCAGACTTCGAACTTCAGAAATTTGTCAG 6735

Qy 1967 TATAAAAAATTTCTCAGTTTCAAAAGAAATGGAACAATATATTTAAACAATCAGAAGTAGTT 2026  
Db 6736 TGGTCAAAAATTTCTCTCATATGATGATGAACTCTTACATGAAGAAGCTGAGATTGTT 6795  
Qy 2027 ATTTGCGCAGGAGCCCGCTACTTTTATGAATTCATTTATCCAAAGGAAAAACAATTA 2086  
Db 6796 ATCACACATGCGCGCCAGCGGTTTATGTCTAGTTATTTCTTTAGGGAATTTACCAAGTT 6855  
Qy 2087 TTGTTTCTTAGACAACAAAAAGTATGCTGAACATGTAATGATCATCAAGTAGAGTTTGTGA 2146  
Db 6856 GTTGTCTTAGGAGAAAGCAGTTTGTGAACATATCAATGATCATCAAAATACAAATTTTGA 6915  
Qy 2147 AGAAGAAATTTTACAAGATTAATATATTTTATTTATAGAAAATATAGATGATTTGTTGAA 2206  
Db 6916 AAAAAAATTTGCCACCTGTATCCCTTGGCTTGGATTTGAAGATGTAGATGACTTTGCGGA 6975  
Qy 2207 AAAATTTATGAAGTTTCTAAGCAAACTAACTTTTACATCAATTAATTTTCTGTGAA 2266  
Db 6976 GCGTTGAAAGGAATATAGCTACAGAAAATATCAGGGAATTAATGATATGTTTGTGTCAT 7035  
Qy 2267 AGATTAAAAACAATAGTTTGAATAATTT 2292  
Db 7036 AAATTAGAAAAAATTTATAGTGAAT 7061

## RESULT 6

US-09-751-708A-83  
; Sequence 83, Application US/09751708A  
; Publication No. US20030157113A1  
; GENERAL INFORMATION:  
; APPLICANT: TERMAN, David S  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE  
; FILE REFERENCE: 751708  
; CURRENT APPLICATION NUMBER: US/09/751,708A  
; CURRENT FILING DATE: 2002-10-15  
; PRIOR APPLICATION NUMBER: US 60/173,371  
; PRIOR FILING DATE: 1999-12-28  
; NUMBER OF SEQ ID NOS: 166  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 83  
; LENGTH: 17276  
; TYPE: DNA  
; ORGANISM: Streptococcus agalactiae  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (7062)..(8207)  
; OTHER INFORMATION:  
US-09-751-708A-83

Query Match 10.4%; Score 726.8; DB 10; Length 17276;  
Best Local Similarity 59.2%; Pred. No. 1.7e-92;  
Matches 1366; Conservative 0; Mismatches 902; Indels 38; Gaps 6;

Qy 16 TGGCATATTTGATATGATAGCAGTTGCGAATTTCTGCAATCTTAAACAAGTCATATACCAA 75  
Db 4765 TAGCGATGATTCAAACAGTTGTAGTTTATTTTCTGCAAGTTTGAACATTAATTA 4824  
Qy 76 ATGCTCATTTAAATCG---TTCCTGGAATTTTATCATATATGATGTTTCATTATTTGTCAT 132  
Db 4825 CTCCTCAATTTTAAACAATTAAGATTTATTTGTTCTTATTTGATACATATATTTGTTT 4884  
Qy 133 TTTTATATCTCGTATGCCAGTTTGAATTTTGAATATAGAGTAATCTGATAGAGTTTGAAA 192  
Db 4885 TTATCTTCTGATTTTACAGAGACTTTTGGAGTCGTGGCTATCTTGAAGAGTTTAAAA 4944  
Qy 193 AAACATTTTAACTATAGTATTAATTTTGGCAATTTTCTTACGGCAGTATCATTTTGTGG 252  
Db 4945 TGGTATTTGAATACAGCTTTTACTATATTTTTCATATATCAAGTTTCATTATTTTATTTTA 5004  
Qy 253 AGAATAATTTCCGACCTTCAAGACGTTGGTGGCTGTATTTTACATTAATTAACATTCGTT 312  
Db 5005 AAAACTCTTTTACACGACGACCTTTCCTTTTCTTATTTGCTATGAAATTCGATTT 5064

Qy	313	TGGTATACCTATTTAAACGTTAAATTATTAAGCAGTGTAAAGATAGCTTCTATATTTTCGACAA	372
Db	5065		
Db		TATTATATCTATTGAAATTCATTTTAAATAATATATCGAAATAATTCCTTAGCTAAGTTTTT	5124
Qy	373	TCTATCAAAAAAGACGATTTCAATTATCAACGGCTGAACGATGGGAAAAATATCGAAGTTT	432
Db	5125	CACGAGATACCAAGTGTGTTTGATACGAATGAAGTATCTTTATCAAAATGACCTTTA	5184
Qy	433	TATTTGAATCACATAAACAATTCAAAAAATCTTGTGTCATTTGGTAGTTTATAGGTACAG	492
Db	5185	GGAAATAAAT---ACGACCATAAATATATCGCTGTCTGTATCTTGGAGCTCCTCTGAAAAGG	5241
Qy	493	AAATAGATAAAATTAATTTATCATACCGCTCTATATTTCTGTGGGAAGACTATAGAGT	552
Db	5242	ATTGTTATGATTTGAAAACATAAATCGTTAAGGATAATAAACAAGATGCTCTTACTTTCAG	5301
Qy	553	TTTCAAACAAGGGAAGTGGTCGACACACGCTCTTTATAAATCTACCAAGTGAAGTTTTAGACG	612
Db	5302	AGTTAACTGCTTAACTGTTGATCAAGCTTTTATTAACATACCCATGATTAATTTGGTA	5361
Qy	613	TAAAGCAA-----TTCGTTTCAGATTTTGAGTTGTTAGTATTTGATGTAAAGCGTTG	663
Db	5362	AATACCAAAATACAAGATATTAATTAATGACATTTGAAGCAATGGGAGTGATTTGTCAAATGTTA	5421
Qy	664	ATATTAAATTCATTCGGTTTTACTCGGTTGAAAACAAAAAATCCAACTGCTAGGTGACC	723
Db	5422	ATGTAGAGGCATTAGCTTTTGATAATAATAGGAGAAAGCGAATCCAAACATTTTGAAGGAT	5481
Qy	724	ATAGCATTTGAACTTTTTCACAAAATTTTATAAGCCTAGTCTCATATCATGATCAAAACGAC	783
Db	5482	ATAGTGTATTACATATTTCTATGAAATTTCTATAATATAGTCCACCTATATAGCAAAACGAT	5541
Qy	784	TTTTGGATATCTCGGACGGTAGTCGGGTTAAATTAATTTGTGGTATAGTTTCTATTTTGT	843
Db	5542	TTTTGGATATCACGGGTGCTAATTATAGTTTGTCTCATATGTGGCAATTTGTGGCAATTTTC	5601
Qy	844	TAGTTCCAAATTTATTCGTAGAGATGGTGGACCGGCTAATTTTTGCTTCAGAAACGAGTTGGAC	903
Db	5602	TAGTTCCGCAAAATCAGAAAAGATGGTGGACCGGCTATCTTTCTCAAAATAGAGTAGGTC	5661
Qy	904	AGAATGGACGCATATTTACATCTTACAAGTTTCGATCGATGTATGTTGATGCTGAGGAGC	963
Db	5662	GTAATGTTAGGATTTTTTAGATTTCTATAAATTCAGATCAATTCGAGTAGATGCAGAACAA	5721
Qy	964	GCAAAAAGACTTGCTCAGCCAAAACAAGATGCAGGGTGGGTATGTTTAAATCGGAA	1023
Db	5722	TTAAGAAGAATTAATTAGTTTCAAAATCAAAATGACGGG-----CTAATGTTTAAAGTTAGA	5776
Qy	1024	AAACGATCCTAGAAATTAATCTCAATGAGCAATTTCAATCGCAAAAACAAGTTTATAGACGAGT	1083
Db	5777	AGATGATCCTTAGAAATTAATAAATAGGAAATTTTATTCG-AAAAACAAGCATAGATGAGT	5835
Qy	1084	TACCACAGTTTTTATAATGTTTTAAATTTGGCGATATGAGTCTAGTTGGTACACGTCACCTTA	1143
Db	5836	TGCCTCAATTTCTATAATGTTTTAAAGGCGATATGAGTTTAGCAGGAAACAGCCCTCCCA	5895
Qy	1144	CAGTTGATGAATTTGAAAAATATATCTCCTGGTCAAAGAGACGATTTAGGTTTTAAACCAG	1203
Db	5896	CAGTTGATGAATATGAAAAGTATAATTAACACGCAAGCGCACGCTTAGTTTTTAAGCCAG	5955
Qy	1204	GGATTACAGGTCCTGCGAGGTTAGTGGTCGTAGTAAATATACACAGACTTCGACGACGATAG	1263
Db	5956	GAATCACTGGTTTGTGGCAAAATATCTGGTAGAAATATATATCTGATTTTGTATGAATCG	6015
Qy	1264	TTCCGTTGGACTTAGCATACATTTGATTAATTTGGACTATCTGGTCAGATATTTAAATTTTAT	1323
Db	6016	TAAAGTTTAGATGTTTCAATATATCAATGAATGGTCTAATTCGTTCAGATATTTAAGATTATTC	6075
Qy	1324	TAAAGACAGTGAAGTTGTATTTGTTGAGAGGGAAGTAGTAAAGTA-----	1372
Db	6076	TCCTAACACATTAAGGTAGTTTAACTCCGGACAGGAGCTAAGTAAGGTAGGTTTCAAAG	6135

## RESULT 7

RESULT /  
US-10-428-817A-79

US-10-428-817A-79  
: Sequence 79, Application US/10428817A; sequence 13, Application US/104  
; Publication No. US20040214783A1

: PUBLICATION NO: US20

; GENERAL INFORMATION:  
; APPLICANT: TERMAN, David S

APPLICANT: LERMAN, DAVID S  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE

FILE REFERENCE: 38373-189118

```
; CURRENT APPLICATION NUMBER: US/10/428,817A
; CURRENT FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: US 60/378,988
; PRIOR FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: US 60/389,366
; PRIOR FILING DATE: 2002-06-15
; PRIOR APPLICATION NUMBER: US 60/406,697
; PRIOR FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US 60/406,750
; PRIOR FILING DATE: 2002-08-29
; PRIOR APPLICATION NUMBER: US 60/415,310
; PRIOR FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: US 60/415,400
; PRIOR FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: US 60/438,686
; PRIOR FILING DATE: 2003-01-09
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 79
; LENGTH: 17276
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7062)..(8207)
US-10-428-817A-79

Query Match      10.4%; Score 726.8; DB 18; Length 17276;
Best Local Similarity 59.2%; Pred. No. 1.7e-92;
Matches 1366; Conservative 0; Mismatches 902; Indels 38; Gaps 6;

Qy 16 TGGCAATTATTGATATGATAGCAGTTCGAATTTCTGCAATCTTAAACAGTCATATACCAA 75
Db 4765 TAGCGATGATTCAAACAGTTGTAGTTATTATTTCTGCAAGTTTGACATTAACATTAATTA 4824

Qy 76 ATGTCGATTTTAAATCG--TTCTGGAATTTTATCATAAATGATGGTTCATTATTTTCAT 132
Db 4825 CTCCCAATTTTAAAGCAATAAAGATTTATTTGTTGTTCTATTGATACATATATTTGTT 4884

Qy 133 TTTTATATCTCGATGCGAGTTGAAATTTGAGTATAGAGGTAATCTGATAGAGTTTGA 192
Db 4885 TTTATCTTCTGATTTTACAGAGACTTTTGGAGTCGTGGCTATCTTGAAGAGTTTAA 4944

Qy 193 AAACATTTAACTATAGTATATATTTGCAATTTTCTTACGCGAGTATCATATTTTGTGG 252
Db 4945 TGGTATTGAAATACAGCTTTTACTATATTTTTCATATCAAGTTCATATTTTATTTT 5004

Qy 253 AGAATAATTTTCGCACTTTCAAGACGTGTGCGCGTATTTTACATTAATAAATTCGTT 312
Db 5005 AAACCTCTTTTACAGCAGCACTTTTCCTTTTATCTTTTATTTGCTATGAAATTCGAT 5064

Qy 313 TGGTATACCTATTTTAACTAATTAATTAAGCAGTTTAAAGGATAGCTTTCTATTTTCGCAA 372
Db 5065 TATTATATCTATTGAAATTCATTTTAAATAATTTATCGAAATAATTTCTTACGCTAAGTT 5124

Qy 373 TCTATCAAAAAGACAGATCTTAATTAACACGGCTGACAGTGGGAAATATGCAAGTTT 432
Db 5125 CACGAGATACCAAGTTGTTTGTGATAACGATAAGGATTTCTTTATCAAAAATGACCTTT 5184

Qy 433 TATTTGATACATAAACAATTCAAAATAATCTGTTGCAATGCTAGTTTGTAGGTACAG 492
Db 5185 GGAATAAAT---ACGACCAATTAATATTCGCTGTCTGTATCTTGGACTCTCTGAAAGG 5241

Qy 493 AAATAGATAAAATTAATTTATCATTCGCTCTATTTATTTCTGTGGAAAGAGCTATAGAGT 552
Db 5242 ATTGTTATGATTTGAAACATAAATCGTTAAGGATAATAAACAAGATGCTCTTACTTCAG 5301

Qy 553 TTTCAACAGGAAGTGTGACACAGCTCTTTATAATCTTACCAAGTGAGTTTGTAGACG 612
Db 5302 AGTTAACTGCTTAACTGTTGATCAAGCTTTTATTAACATACATCCCAATTTGAATTTTGGTA 5361

Qy 613 TAAAGCAA-----TTCTGTTTCAGATTTTGAAGTTGTTAGGTATTGATGCTAAGCGTTG 663
Db 6436 TGGTGTGAAGACCGTTTATATAGAGGTTTTCGACAGGATAGATAAACAACACTTTTGACAGG 6495
```

```
Db 5362 AATACCAATACAGATATATTATTAATGACATTTGAAGCAATGGGAGTGATTTGTCATGTTA 5421
Qy 664 ATATTAAATCATTCGGTTTTTACTGCGTTGAAAAACAAAAATCCAACTGCTAGGTGACC 723
Db 5422 ATGTAGAGGCACCTAGCTTTGATAATATAGGAGAAAAAGCAATCCAACTTTTGAAGGAT 5481
Qy 724 ATAGCATTTGTAACCTTTTCCACAAATTTTATAGCCTAGTCTATATCATCATGATGAACGAC 783
Db 5482 ATAGTGTATTACATATTCTATGAAATTTCTATAAATATATAGTCACTTTATAGCAAAACGAT 5541
Qy 784 TTTTGATATACCTCGGAGCGGTAGTCGGGTAAATTTTGTGTATAGTTTCTATTTTGT 843
Db 5542 TTTTGGATATCAGGGTGTCTATTATAGTTTGTCTATATGTGSCATTTGTGGCAATTTTTC 5601
Qy 844 TAGTTCCAAATTTCTGTAGAGATGGTGGACCGGCTATTTTGTCTAGAAAACGATTTGGAC 903
Db 5602 TAGTTCGCAAAATCAGAAAAGATGGTGGACCGGCTATCTTTCTCAAAATAGAGTAGGTGC 5661
Qy 904 AGAATGACGACATATTACATCTCAAGTTTTCGATCGATGTATGTTGATGCTGAGGAGC 963
Db 5662 GTAATGGTAGGATTTTGTAGATTTCTATAAATTCAGATCAATGCGAGTAGATGCGAAACAA 5721
Qy 964 GCAAAAAGACTTGTCTCAGCCAAAACAGATGCAAGGTGGGTATGTTTAAAAATGGAA 1023
Db 5722 TTAAGAAAGATTTTATAGTTTCACAATCAATGACAGGG-----CTAATGTTTAAAGTTAGA 5776
Qy 1024 AAACGATCCTAGAAATTTACTCCAATTTGACATTTTATACGCAAAAACAAAGTTTAGACGAT 1083
Db 5777 AGATGATCCTAGAAATTTACTAAAATAGCAAAATTTATTTCG-AAAAACAAGCATAGATGAGT 5835
Qy 1084 TACCACAGTTTATATGTTTAAATTTGGGATATGATCTAGTTGGTACAGCTCCACCTA 1143
Db 5836 TGCCTCAATTTCTATAATGTTTAAAGGCGATATGATTTAGCAGGAACACGCGCTCCCA 5895
Qy 1144 CAGTTCATGAATTTGAAAAATATACCTCGGTCAAAAGACAGCATTTAGTTTAAACCAG 1203
Db 5896 CAGTTGATGAATATGAAGATTAATTAACGCGAAGACGCGCTTAGTTTAAAGCAG 5955
Qy 1204 GGATTAACAGTCTCTCGCAGGTTAGTGGTGTAGTAAATATACAGACTTTCAGCAGCTAG 1263
Db 5956 GAATCACTGGTTTGTGCAAAATATCTGGTAGAAATAATATTACTGATTTTGTAGAAATCG 6015
Qy 1264 TTCGTTGGACTTAGCATATCATTTGATTAATTTGGACTATCTGGTACAGATATTTAAATTTAT 1323
Db 6016 TAAAGTTAGATGTTCAATATATCAATGAAATGGTCTATTTGGTCAGATATTAAGATATTC 6075
Qy 1324 TAAAGACAGTGAAGATTTGTATTGTCAGAGAGGGAAGTAAAGTAAAGTAAAGTAAAGTAAAG 1372
Db 6076 TCCTAACACTAAAGGTAGTTTACTCGGACAGAGCTAAGTAAAGTAAAGTAAAGTAAAGTAAAG 6135
Qy 1373 -----TATGAAAGTTTGTGGTTCGTTTTCAGGGGACATTTGACTCACTTGTATTT 1426
Db 6136 GAATAATAATGAAATTTTGTCTGGTTGGTTCAAGTGGTGGTCACTAGCACACTTGAACCT 6195
Qy 1427 GTTAAACCGTTTTCGAAAGGAAGAACGTTTTTGGTTAAACATTTTATGATAAAGAGATGC 1486
Db 6196 TTTGAAACCCATTTTGGAAAAAAGATAGTTTGGGTAACTTTTGTATTAAGAGATGC 6255
Qy 1487 AAGAAAGTCTTTTGAAGATGAAATAATGTATCCATGTTACTTTTCCAAACAAATCGCAATCT 1546
Db 6256 TAGAGATTTCTAAGAGAGAGATTTGTATATCATTTGTTCTTTTCCAAACAAACCGTAAATGT 6315
Qy 1547 CATTAATTTAGTGAATAATACTTTCTTAGCTTTTCAAAATTTTACGTGATGAGAAACGAGA 1606
Db 6316 CAAAAACTTGGTAAAAATACTATTCTAGCTTTTAAAGTCTTTAGAAAAAAGAAAGACGAGA 6375
Qy 1607 TGTATATTATCTGCTGGTCCGCGCTTCTGTCCTTCTTTTATCATCGGAAACATTT 1666
Db 6376 TGTATATCATATCTGCTGGTCCGCTGTAGCAGTACCAATCTTTTATATTTAGTAAATTT 6435
Qy 1667 TGGAGCAAGACCATTTATATTTGAAAGTATTTGATCGAGTTTAAATAATCTCATTTAACTGG 1726
Db 6436 TGGTGTGAAGACCGTTTATATAGAGGTTTTCGACAGGATAGATAAACAACACTTTTGACAGG 6495
```

```
Qy 1727 AAAAAGTATATCCCGTAAACAGATATATTTTATTTGTTTCAGTGGGAGAAATGAAAGAGGT 1786
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 6496 AAAATTAGTGTATCTGTAAACAGATAAATTTATTTGTTTCAGTGGGAGAAATGAAAGAGGT 6555
Qy 1787 ATATCTAAATCTATTAATCTGGGAGATATTTTATTTGTTTCAGTGGGAGAAATGAAAGAGGT 1846
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 6556 TTATCTTAAGCAATTAATTTAGAGGAAATTTTATTTGTTTCAGTGGGAGAAATGAAAGAGGT 6615
Qy 1847 CATGAACAACAGTTTAAATCGATTGATAAAGAGATTTGATTTTATTTGAAAGAAATGGAAGT 1906
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 6616 CATGAACAGCAGTTCAACCGTCTTATTAAGAGATTTGATAGATTAAAGAGGACAGGTGCT 6675
Qy 1907 ATACCCAGCAGAAATTTATTTCAACAGGATTTCTGACTATATTTCCAGATATTTGCAAG 1966
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 6676 ATTGATCAAGAGAGTGTTCATTTCAACCGGTTTACTCAGACTTCGAACCTTCAGATTGTCAG 6735
Qy 1967 TATAAAAAATTTTCAGTTTCAAGATTAAGAAAGATGAACAATATATTTAAACAATCAGAAGTAGTT 2026
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 6736 TGGTCAAAATTTCTCTCATATGATGATGAATGAATCTTTACATGAAGAGCTGAGATTGTT 6795
Qy 2027 ATTTGCCACGAGGCCCGCTACTTTTATGAATTCATTTATCCAAAGGAGAAACAAATTA 2086
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 6796 ATCACAATGCGCGCCAGCGCTTATGTCAGTTATTTCTTTAGGGAATTTACCAGTT 6855
Qy 2087 TTGTTCTTAGCAAAAAAGTATGGTGAACATGTAATGATCATCAAGTATAGATTGTA 2146
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 6856 GTTGTCTTAGGAGAAAGCAGTTTGGTGAACATATCAATGATCATCAAAATCAAAATTTTA 6915
Qy 2147 AGAAGATTTTACAAGATTAATATTTATTTATGAGAAATATAGATTTGTTGAA 2206
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 6916 AAAAAAATTTGCCACCTGTATCCCTTGGCTTGGATTAAGATGATGAGTGGACTTGGCGAA 6975
Qy 2207 AAAATTTAAGATTTCTAAGCAAACTAACTTTATCATCAAAATTAATTTTGTGAA 2266
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 6976 GCGTTGAAAGGAATATAGCTACAGAAATATCAGGGAATATGATGTTTGTAT 7035
Qy 2267 AGATTAAACAAATAGTTGAAATTT 2292
Dy 7036 AAATTAGAAAAAATTATAGTGAAT 7061

RESULT 8
US-10-192-280-1
; Sequence 1, Application US/10192280
; Publication No. US20040009574A1
; GENERAL INFORMATION:
; APPLICANT: Dattagupta, Nanibhushan
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
; TITLE OF INVENTION: STREPTOCOCCUS AGALACTIAE CAPSULAR POLYSACCHARIDE SYNTHESIS
; FILE OF INVENTION: GENES
; FILE REFERENCE: 475412001300
; CURRENT APPLICATION NUMBER: US/10/192,280
; CURRENT FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 25020
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
US-10-192-280-1

Query Match 10.3%; Score 721.4; DB 17; Length 25020;
Best Local Similarity 58.9%; Pred. No. 1.1e-91;
Matches 1389; Conservative 0; Mismatches 926; Indels 42; Gaps 7;

Qy 16 TGGCATTATTTGATATAGTAGCAGTTGCAATTTCTGCAATCTTAAACAGTTCATATACCA 75
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 5154 TAGCATGATTTCAACAGTTGTGGTTTATTTTCTGCAAGTTGATCAATTAATTA 5213
Qy 76 ATGCTGATTTAAATCGTTCTGGAAATTTT-----TATCATTAATGATGGTTTCAATTTTGCA 131
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Dy 5214 CTCCCAATTTTAAAGCAATAAGATTTATTTGTTTCTAATTTGATACATTATATTGTC 5273
Qy 132 TTTTATATCTCGTATGCCAGTTGAAATTTGAGTATAGAGTAAATCTGATAGAGTTGAA 191
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 5274 TTTTATCTTCTGATTTTACAGAGACTTTTGGAGTCTGGCTATCTTGAAGAGTTTAA 5333
Qy 192 AAAACATTTAACTACTAGTATAATTTTCAATTTTCTTACGGCAGTATCATTTTGTG 251
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 5334 ATGGTATTGAATATACAGCTTTTACTATATTTTCAATCAAGTTCAATTTTATTTT 5393
Qy 252 GAGATAATTTTCGCACCTTTCAAGACGCTGGTCCCGTGTATTTTCAATTAATAAATCTCGTT 311
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 5394 AAAAATCAATTTACACGACACGACTTTCTTTTCTTTTATTTGCTATGAATTCGATT 5453
Qy 312 TTGGTATACCTATTATTAAGTAAATTTAAGCAGTTTAAAGATAGCTTTCTATTTTCGACA 371
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 5454 TTATTTGATCTTATTTGAATTCATTTTAAATAATTTATCGAAAAATATTTCTTACGCTAAAGTTT 5513
Qy 372 ATCTATCAAAAAAGACGATTTCTAATTTACACGCTGAAACGATGGGAAAAATATCAAGTT 431
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 5514 TCACGAGATACCAAGTTGTTTGTATACGAATTAAGGATTTCTTTATCAAAAATGACCTTT 5573
Qy 432 TTATTTCAATCACATAAACAATTTCAAAAAATCTTGTTCATTTGGTAGTTTATAGGTACA 491
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 5574 AGGATTAAT---ACGACCATTAATTTATTTGCTCTGTCATCTTGGACTCTCTGAAAAAG 5630
Qy 492 GAAATAGATAAAATTAATTTATCAATACCGCTCTATTATTCTTGTGGAAGAGCTATAGAG 551
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 5631 GATTGTTATGATTTGAAAAACATACTCGTTAAGGATAATTAACAAAGATGCTCTTACTTCA 5690
Qy 552 TTTTCAACAAGGAGTGGTGCACACGCTCTTTATAAATCTACCAAGTGAATTTTAGAC 611
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 5691 GAGTTAACTGCTTAACTGTTGATCAAGCGTTTATTAACATATACCATTGAAATTTTGGT 5750
Qy 612 GTAAGCAATTCG-----TTTCAGATTTTCAGTTGTTAGTATTGATGTAAGCGTT 662
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 5751 AAATCCAAATACAGATATTTAATTAAGCATTTGAGCAATGGAGTGAATTTGCAATTT 5810
Qy 663 GATATTAATTCATTCGGTTTTTACTCGCTGTAAGAAACAAAAAATCCAACTGCTAGGTGAC 722
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 5811 AATGTAGAGGCATTTAGCTTTGATAATATAGGAGAAAGCGAATCCAAACTTTTGAAGGA 5870
Qy 723 CATAGCATGTGTAATTTTCCAAATTTTATTAAGCTTAGTCAATATCATGATGAAGGA 782
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 5871 TATAGTGTATTATCATATTTCTATGAAATTTCTATAAATATAGTCACTTTATAGCAAAACGA 5930
Qy 783 CTTTGTGATATCTCGGAGCGGTAGTGGGTTAATTTATTTGTTGTTAGTTTCTTATTTTG 842
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 5931 TTTTGTGATATCAGGCGTCTATTATAGTTTGTCTATATGTCATTTGGCATTTGGCAATTTT 5990
Qy 843 TTAGTTCCAAATTTCTGTAGAGATGGTGGACCGCTATTTTGTCTCAGAAACGAGTTGGA 902
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 5991 CTAGTTCCCAAAATCAGAAAGATGGTGGACCGCTATCTTTTCTCAAAATAGAGTAGGT 6050
Qy 903 CAGAAATGAGCATATTTATCAATTTCAAGTTTCGATCGATGATGTTGATGCTGAGAG 962
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 6051 CGTAAATGTTAGGATTTTATAGATTTCTATAAATTCAGATCAATCGAGTAGATGCGAAACAA 6110
Qy 963 CGCAAAAAGACTTGTCTCAGCCAAAACAGATCAAGGCTGGGTATGTTTAAATGGA 1022
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 6111 ATTAAGAAAGATTTATTTAGTTTCAAAATCAAGACAGG-----CTAATGTTTAAAGTTAG 6165
Qy 1023 AAAACGATCTTAGAATTTACTCCAATTTGACATTTTATACGCAAAAAAACAAGTTTATAGACGAG 1082
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 6166 ACGATGATCTTAGAATTTACTAAAATAGAAAATTTATTCG-AAAAACAGCATAGATGAG 6224
Qy 1083 TTACCACAGTTTATAATTTTAAATTTGGCGATATGATCTAGTTGGTACAGTCCACT 1142
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 6225 TTGCCTCAATTTCTATATGTTTAAAGGTGATGATGATGTTTATAGTAGGAACACGCGCTCCC 6284
Qy 1143 ACAGTTGATGATTTGAAAAATATATCTCTGCTCAAAAGAGACGATGAGTTTAAACCA 1202
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 6285 ACAGTTGATGATTAAGAAAGTATAATTTCAACGAGAGACGCGCTTTAGTTTAAAGCCA 6344
```



Db 6675 GTTTTGAATTTTACAAATTTTGAAGCATGTACACCGATGCCGAGAACGCCAAAGAA 6734  
Qy 975 TTGCTCAGCCAAACACGATGCAAGGGTGGGTATGTTTTTAAATGGGAAACGATCTTA 1034  
Db 6735 TTACTAAACACAAATGATTTGGATACCTGACTTAATGTTTGAAGTGG--ATGATGACCCCTC 6792  
Qy 1035 GAATTTACTCCAAATGGCAATTTATACGCAAAACAAAGTTTGAAGAGTTTACCAAGTTT 1094  
Db 6793 GTATCTTCCCAATTTGGACATAAGTTTACGTGATTTGGTCACTT-GATGAATTTACCAATTT 6851  
Qy 1095 TATAATGTTTTAAATTTGGCGATATGCTGTAGTTGTACACGTTCCACCTTACAGTTGATGAA 1154  
Db 6852 ATTAATGTCCTAAAGGTGAATGTCGTGTGGGCACAGTCCACCAACCGCTTGACGAA 6911  
Qy 1155 TTGAAAATATATCTCTGGTCAAAAGAGACGATTTAGTTTTTAAACACGAGGATTTACAGGT 1214  
Db 6912 TATCATCTATGAGTTACATCACTTCAACCGATTTGACACCAACCAAGGAATTTACTGTT 6971  
Qy 1215 CTCTGGCAGGTTAGTGGTCGTAGTAATATACAGACTTCGAGGAGCTAGTTGGTTGGAC 1274  
Db 6972 TTATGGCAAGTTAGCGGTCGTAGTGACATTTACCGACTTTGAAGAAGTCGTAGCACTTGAT 7031  
Qy 1275 TTAGCATACATTTGATTTGACATCTGCTCGATATTAATTAATTTTAAAGACAGTG 1334  
Db 7032 ATGAAGTATATCCAAACCTGGAGCATCAGTGAAGATATTAATAATTTATGCCCCAAACATTT 7091  
Qy 1335 AAAGTTGATTTGTGAGAGAGGAAGTAAGTAA--AGTATATGAAGTTTGTGTTGGTCG 1392  
Db 7092 GGAGTCGTACTAAAGAGAGGGAAGTAAGTAGATATATTTATGNAAGTTTGTGTTAGTAG 7151  
Qy 1393 GTTCTCAGGGGACATTTGACTCACTTGTATTTGTTTAAACCCGTTTTTGGGAAGGAAG 1452  
Db 7152 GTTCTCTCGTGGACATTTGGCACATTTGAAATATGCTAAACCCCTTTTGGAGTGAACATA 7211  
Qy 1453 AACGTTTTTGGGTAACTTTGATTAAGAGGATGCAAGAGCTTTTGAAGATGAAGAA 1512  
Db 7212 GCCGTTTCCGGGTACATTTGATTAAGAGAGCGCAAGAGTGTGTTAAGTCATGAAGAAAT 7271  
Qy 1513 TGATTCATGTTACTTTCCAAACAAATCGCAATCTCATTAATTTAGTGAAGAAATACTTTCT 1572  
Db 7272 TTTATCCGTGTTATTTCCGACTAACAGAAATTTAAGAAATTTGGTAAGAACACTTTCT 7331  
Qy 1573 TAGCTTTCAAATTTTACGTGATGAGAAACAGATGTTTATTTATTTCACTCGTGGCGCG 1632  
Db 7332 TAGCACTTGAATTTTAAAGAAAGAAACCTGACGTTTATTTATTTCACTAGGAGCAGCG 7391  
Qy 1633 TTGCTGCCCTTCTTTTACATCGGAAACATTTTGGAGCAAGACGATTTATTTGAAG 1692  
Db 7392 TAGCAGTTTCCATTTTATCTGGGTAACTGTTTGGAGCGAAACCGTTTTATATCGAAG 7451  
Qy 1693 TATTTGATCGAGTTAATAAATCTACATTAACCTGGAAACACTGTTTATCCCGTAAACAGATA 1752  
Db 7452 TATTTGATAGAAATAGATAAACCGCTGTGACTGGAAAGTTGGTTTATCCAGTGACAGATA 7511  
Qy 1753 TTTTATTTGTTTCACTGGGAAGAAATGAAGAGGTATATCTTAAATCTATTAACTTTGGGGA 1812  
Db 7512 AATTTATTTGTTTCACTGGGAGGAGTAAACCTGCTATCCCAAGCTATTATCTGGGGA 7571  
Qy 1813 GTATTTTAAATGATTTTGTAAACAGTAGGAACTCATGAACAAACAGTTTAAATCGATGAT 1872  
Db 7572 GTATTTTAAATGATTTTGTATACAGTTTGAACCTCATGAACAGCCCTTTTAAATAGGCTTAT 7631  
Qy 1873 AAAGAGATTCGATTTATTTGAAGAAATTTGAAGTATTAACCGNAGAAATTTTATTTCAAC 1932  
Db 7632 TAAGGAAGTTGATTCGTTTTTAAAGAAAGAGGTATTATTAACAGATGAGGTTTTTATTCAGAC 7691  
Qy 1933 AGGATATCTGACTATATTTCCAGATATGCAAGTATATAAATAATTTCTCAGTTACAAAGA 1992  
Db 7692 AGTTTTTCAACTATATGACCTCAATCTAGTCACTGGAAATAATATTTCTTATTTCTGA 7751  
Qy 1993 AATGAAACATATATTAACAAATACAGAGTAGTTATTTTCCACGAGGCCCGCTACTTTT 2052  
Db 7752 AATGGAAGATTACATGAATCGTGACAGATATTTATATACCGCATGTTGGTCCAGCGACATTT 7811

Qy 2053 TATGAATTCATTTATCCAAAGGAAAAACAAATTTATGTTTCTCTAGCAAAAAAGTATGG 2112  
Db 7812 CATGGGCAATTCGTTAAAGGAAAAAACCCGATGTTGTTCCAGACAGGAAAAAGTTGG 7871  
Qy 2113 TGAACATGTAATGATCATCAAGTAGAGTTTGTGAAGAATTT---TACAAGATAATAA 2169  
Db 7872 AAGCATGTAATGATCATCAAGTTGAGTTTGTGGAACAGGTTCTGACCGATTTGGAG 7931  
Qy 2170 TATTTATTTATAGAAAATATAGATGATTTGTTTGAAGAAATTTATGAGTTTCTTAAGCA 2229  
Db 7932 TATCGTTGCTGAGAGAAATTAATGATTCGAAATTTATTTAATTTAGATTTAATTTGT 7991  
Qy 2230 AACTAACTTTACATCAAAATAATATTTTGTGAAAGATTTAAACAAATAGTTGAAAA 2289  
Db 7992 AGATGAAAGTTTCCAAATTCGAAACCACTAAGATTTAATAGTCAATTTAAACCAAG---AAA 8047  
Qy 2290 ATTTAATGAGATCAAGAAATGAATTAATAAAGATGATATTTGATAATGGCTTATC 2349  
Db 8048 TAGAAAGTTTGGTTAGATGAATGATTCCTAAAGATTCATTTATTTGGTTGGAGAA 8107  
Qy 2350 ATAAATTTTCTCAGATTTTACTCGAGAGGGATACAGATATTTATCATCTCTCTCAGGAGA 2409  
Db 8108 ATCCTCTTCTGACAGTGTAAAAATTTGTATAAATTCGTGGAATAAATCTGTCGAAAT 8167  
Qy 2410 ATGCACACCAATTTAGTTCCCTCAGAAATACCTGTATAATTTATTTTAAAT-ATTCTCAGGAT 2468  
Db 8168 ATGAATAATCGAATCGAATCGAATCAAAATTAATGATGATACATAAAATTCATATTTCTG 8227  
Qy 2469 TATATGTTGAATTTTACAAAGATGAGCAAAATATATAAGAAATAGGATATATGAAACAG 2528  
Db 8228 AAGCTTATAAATAAAGAAATATGCTTTGTTATCTGACTATGCTAGGCTAGATATCATAT 8287  
Qy 2529 TTAATGTTTACAGATTTTCTCCTAATA-----TATCAGAAAAAACTATTGATAATGT 2580  
Db 8288 ATATAGGGCGGTTTTTATTTAGATCTGATGTTGAATTTGTTAAGCATTTGGACGAT 8347  
Qy 2581 ACTGTTTAGAATTTTATTAAGAATGTATCGAGCTTTTGAATACTATTATTACAAAGATTGT 2640  
Db 8348 TAACTCTCGAACACTGTTATATGGAATGGAACCAAGTGGTGGTCTGTTAATCTCGATTAG 8407  
Qy 2641 GTTTATGATAGATAAAAAACATGGTCTAAGAAATAGATTTGGTTCTAATTTGGGTTTCG 2700  
Db 8408 GTTTTGGTGCAGAA-AAAAGGACATCTTTTATAAAGAAAAATATGACGCAATATGAAGAA 8466  
Qy 2701 CTTCACATGATTTTGTGGCAATCTTTTATCAAAATGAAGAAACGAAACAGCTTATTATT 2760  
Db 8467 GTTCTTTTAACTTAAAGTACTAGAAACATGTGTGGATATACGACAAATTTATTATTA 8526  
Qy 2761 AAGTAATCTAAATGTCAGATGAATCTATTATACAGCAATATATAGAAAAATATGAAATTT 2820  
Db 8527 TCAAGGGGTTATTTAGTAGAAATTCATATCAAAAAATTAGTGATGTCTCAATTTATCCA 8586  
Qy 2821 TCAAAATAGATTTCTAAATATGGAATTTAAGATATATAAAGTGGAAAAATCAACATCT 2880  
Db 8587 ACAGATTTTTTTTGTCCGTTTAAATATGCAAAACACAAAGAAATGGAATTAACATAAATACT 8646  
Qy 2881 TCTCCTATTTGCTTACAGATGATTTCTATGATGAATTTGCTAATGCAAGAAATTTAGGT 2940  
Db 8647 TATTCAATTCATCATTTATGATTTCAACTTGGTATG-----GTAATGGTGT 8691  
Qy 2941 TTTTATTTTGTCAAGAGTTAAAAATAGAAATAAATCTAAATTTTAAAGAAATTTATTACT 3000  
Db 8692 AGTCAATTAATAAAGAAATTAATTAACCATTAAGATTAATTCGTATCTTATGAT 8751  
Qy 3001 AAAAAATAAATAGTGTGATTTTGTGAGAGTAAATGATGTTTAAATTTTAAATATGACC 3060  
Db 8752 AATATTTTAGGTGAAGCTCTTATGCTA-AAATCAAGCTATTATTAAGAAATGATAAT 8810  
Qy 3061 CGGAATATTTTATTTTAAAGTACTTCTGGTTGATTTATTTTATCCAGAGCAAAAGTATG 3120  
Db 8811 TTCAAGAGGAGATATTTTGTAACTAATTAATTTGAAATTTTGTGAT-----ATATATATAT 8864





Db 10917 TTGATGATGGATCTACAGACAAATCAATATATCGATATGTAAAGATATGCTGCGATGTGATT 10976  
Qy 5341 AAAGAGTAAAAATTTTTTTCATATCATAGTGGAGTATCAAAATGCTAGAAATCATGAA 5400  
Db 10977 CTCGAATTAAGCTTTTCTAAAGAGATGGCGGTTTCTAGCGCTCGAATCTAGTGC 11036  
Qy 5401 TAAAGCGAGTACAGCTGAATATATATGTTTGTGACTCTGATGATGTTGTTGATGATA 5460  
Db 11037 TTCTACATGTTCAAGGAGAGTACGTTGTGTTGTAGATTCAGATGACTTTGTATCACC 11096  
Qy 5461 GATTAGTAGAAAAATTTATATTTTAAATATATAAAGTAGAAGTGAATTTCTGTTGTT 5520  
Db 11097 AATATTGGAACATTTATCACTTACTATPAAGTACTAAGTCAGATGACTGTTAGCT 11149  
Qy 5521 TGTAGCTACTTTTTTCAGAAAATATAAATAATTTTTGAAAGTGAATCAATCAATATTTGATT 5580  
Db 11150 --TCTGTAAGTCGTTTATAACATTTTGAATAAAGAGTGTAAAGATATCGGATTTATCTT 11207  
Qy 5581 TTGAAGCAATTAATACCGTGCAGGACATGGGAGAAAAAAATTTTATGAAATTTGTATATA 5640  
Db 11208 TTAATCAATAACATCAGATGAAGCTTTAAGAAAAATTTCTTTTAGGTGAGGGGATAAAT 11267  
Qy 5641 ATAATATTTTTTCTACTCTCTGTTGTAACATATATAAGAAAGATACATACAGATCTTT 5700  
Db 11268 GTTATCTTTTTTCAAAAATATTTAAATATGAACATATAAAGGACTCCGATTTGATGAA 11327  
Qy 5701 TTCAAGGAATCAATGGTTAGGAGAAGATTTACTTTTTTAACTCGCATTTTAAAGATA 5760  
Db 11328 GTTTAGATCAGCAGAGACGTTTTGTTTATTTATCAACCTTTAAGAACATAAATTTTG 11387  
Qy 5761 TAGATAGAGTTAGTTATTTGACTGAACATCTTTATTTTATAGGAGAGGTATACATGATA 5820  
Db 11388 CATCTATGATGGCACTGTCAGATTAATTTTATATCTTAA-GAAGAGGATCTTTAACA 11446  
Qy 5821 CAGTAAATCTTTTAAAGAGGTGTTTTTGCATTTGGAAATTTGCAAAAACAGTGA 5880  
Db 11447 AATAAAGACTGACTTCATCAAGAAATGATAGTTCCATTAGAGTTGCGGAATTTATTA 11506  
Qy 5881 TAGTATTTTAAAGCAATATATGCTGAGGATTTGACGTATCAATTTGTTAAAGATAC 5940  
Db 11507 AGAGTTGCAACAGCAAAAAATTTGAAATGTTAAGTGA--AATTAATGAATATCAT 11564  
Qy 5941 TACGTGCGCAAGTATTTTATATAGCTTACTAATGTTTAAATACGGAACACAGCTATTT 6000  
Db 11565 TAAAGGTGAGGTTCTTCTGAGTGGATTTCTTAAATAGTGAACCTTAGAATTCAGTTTGA 11624  
Qy 6001 TTGCAAAATTTTAAATTTTGAATCTTTTAAATAATATATTTTAACTTTGTTAAAG 6060  
Db 11625 AATATTATTAATCATACTGAGAGA-AGTTAGAAGTTTAAATTTGTTACATAAAGTTTCAA 11683  
Qy 6061 TATCTACAAAAATTTCTTTGTTCAAAAAATTTTGTATAAGAAATTTGTTTGAACA- 6114  
Db 11684 TATCTAACTTTTAAAAAATTTTATAGGATTAATTTTAAAGTTAGTCTTAGATAGTT 11743  
Qy 6115 AAGTTTTTAAAAAATATTTATGTTTATATAGGAAGATATCATGATACTATTTAGTAA-- 6172  
Db 11744 ACAATCTTAAAAAATAAATAGTATCTCGGAGGAGTATTCATGATTTTAAATAGTAA 11803  
Qy 6173 ----AATTTCTAATTTGACTATATATATATAGTAAATAATTTATCTAAATGTTATAG 6228  
Db 11804 CTCTTTGTTTCAATTTATTTCAATTTTATAATGTAGAAAAATTTTATAGAACAGTGCTCTA 11863  
Qy 6229 ATAGATTGTAATCAGACCTTACAACA 6256  
Db 11864 CTGAGTGTTCACATAAATTTATGACA 11891

RESULT 10

US-10-804-408-166

; Sequence 166, Application US/10804408

; Publication No. US20040253617A1

; GENERAL INFORMATION:

; APPLICANT: PANRONG, KONG

; APPLICANT: GILBERT, GWENDOLYN  
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI  
; FILE REFERENCE: 675002-2001  
; CURRENT APPLICATION NUMBER: US/10/804,408  
; CURRENT FILING DATE: 2004-03-19  
; PRIOR APPLICATION NUMBER: PCT/AU02/01281  
; PRIOR FILING DATE: 2002-09-18  
; PRIOR APPLICATION NUMBER: AU PR 7749  
; PRIOR FILING DATE: 2001-09-19  
; NUMBER OF SEQ ID NOS: 182  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 166  
; LENGTH: 2226  
; TYPE: DNA  
; ORGANISM: Streptococcus agalactiae  
US-10-804-408-166

Query Match 9.3%; Score 652.4; DB 18; Length 2226;

Best Local Similarity 59.6%; Pred. No. 2.6e-82;

Matches 1225; Conservative 0; Mismatches 791; Indels 38; Gaps 6;

Qy 16 TGGCATTTATTTGATATGATAGCAGTTGCAATTTCTGCAATCTTAAACAAGTCATATACCAA 75  
Db 182 TAGCATGATTCAAACAGTTGTGTTTATTTTCTGCAAGTTTGACATTAACATTAATA 241  
Qy 76 ATGCTGATTTAAATCG--TTCTGGAAATTTTATCATATAATGATGTTTCATTATTTTGCAT 132  
Db 242 CTCCCAATTTTAAAGCAATAAAGATTATTTGTTTGTCTATTGATACATTATATTGTCT 301  
Qy 133 TTTTATATCTCGTATGCGAGTTGAAATTTGATATAGAGTAAATCTGATAGATTTGAAA 192  
Db 302 TTTATCTTCTGATTTTACAGAGACTTTTGGAGTCGTGGCTATCTTGAAGATTTTAAAA 361  
Qy 193 AAACATTTAACTATAGTATATATTTGCAATTTTCTTACGCGAGTATCATTTTGTGG 252  
Db 362 TGGTATTGAATAACAGCTTTTACTATATTTTCAATCAAGTTTCATTATTTTATTTTA 421  
Qy 253 AGAATAATTTTCGCACTTTTCAAGACGTGGTCCGTGTATTTTCAACATTAATAAATTCGTTT 312  
Db 422 AANAATCATTTACATGACAGACTTTTCTTTTCTTTTATTTGCTATGATTCGATTT 481  
Qy 313 TGGTATACCTATTAAACGTAATTAATTAAGCAGTTTAAGGATAGCTTTCTATTTTCGACA 372  
Db 482 TATTATATCTATTGAAATTCATTTTAAATAATATATCGAAAAATTTCTTACGCTAAGTTT 541  
Qy 373 TCTATCAAAAAAGACGATTTCTAATTACACGGCTGACGATGGGAATATGCAAGTTT 432  
Db 542 CACGAGATACCAAGTTGTTTGTATAACGAATAAGGATTTCTTTATCAAAAATGACCTTTA 601  
Qy 433 TATTTGAATCACATAAACAATTTCAAAAAATCTTTTGTCAATTTGTTAGTTAGGTACAG 492  
Db 602 AGAATAAT---ACGACCAATTAATATATCGCTGTCTGTATCTTGGACTCTCTGAAAAG 658  
Qy 493 AAATAGATAAAATTAATTTTATCATTAACGCTCTATTTTCTGTGGAAGAAGCTATAGAGT 552  
Db 659 ATTGTTATGATTTGAAAACATAAATCGTTAAGGATAATAAACAAGATGCTCTTACTTCAG 718  
Qy 553 TTTTCAACGAAGGAGTGTGACCACTCTTTTAAATCTTAAATCTCAAGTGGTTTGTAG--- 609  
Db 719 AGTTAACTGCTTAACCTGTTGATCAAGCTTTTATTAACATACCCATTTGAAATTTTGTGA 778  
Qy 610 -----ACGTAAGCAATTTGTTTTCAGATTTTGTAGTTGTTAGGTATTTGATGTAGCGTTG 663  
Db 779 AATACCAATAACAAGATTAATTAATGACATTGAAGCAATGGAGTGTGTTGTCATGTTA 838  
Qy 664 ATATTAAATTCATTCGGTTTACTCGTTGAAAAAACAATAAATCAACTGCTAGGTGACC 723  
Db 839 ATGTAGAGGCACTTAGCTTTTGATAATATAGGAAAAAGCGAATCCAAACTTTTGAAGAT 898  
Qy 724 ATAGCATTTAACTTTTTCACAAATTTTATAGCCCTAGCTCATATCATGATGAACGAC 783  
Db 899 ATAGTGTATTACATATTTCTATGAATTTCTATAAATATAGTACCTTATACAAACGAT 958



784 TTTTGGATATACGAGCGGTAGTCGGGTTAAATTTTGTGGTATAGTTTCTATTGTT 843  
Db |||||  
959 TTTTGGATATCATGGGTGCTATTATAGTTTGTCTCATATGTGCAATTTTTC 1018  
Qy |||||  
844 TAGTTCAAATTTTCGTAGAGATGGTGGACGGCTATTTTGTCTCAGAAACGAGTTGGAC 903  
Db |||||  
1019 TAGTTCGGCAATCAGAAAAAGATGGTGGACGGCTATCTTTTCTCAAAATAGAGTAGGTC 1078  
Qy |||||  
904 AGAATGACGCATATTTACATTTCTACAGTTTCGATCGATGTATGTTGATCGCTGAGGAGC 963  
Db |||||  
1079 GTAATGTAGGATTTTATAGATTTCTATAATTCAGATCAATGCGAGTAGATGCGAACA 1138  
Qy |||||  
964 GCAAAAAGACTCTCTCAGCCMAAACAGATGCAAGGGTGGGTATGTTTAAAAATGGAA 1023  
Db |||||  
1139 TTAAGAAGATTATTAAGTTTACANTCAATGACAGGG-----CTAATGTTAAGTTAGA 1193  
Qy |||||  
1024 AAACGATCCTAGAAATTAATCCAAATGGACATTTTCATACGCAAAACAAAGTTTACAGAGT 1083  
Db |||||  
1194 CGATGATCCTAGAAATTAATAATAGGAAAAATTTATTCG-AAAAACAAGCATAGATGAGT 1252  
Qy |||||  
1084 TACACAGTTTATTAATGTTTAAATTCGCGATATGAGTCTAGTTGGTACACGTCACCTA 1143  
Db |||||  
1253 TGCCTCAATTTCTAATATGTTTAAAGGTGATATGAGTTTAGTAGGAACACGCCCTCCCA 1312  
Qy |||||  
1144 CAGTTGATGAATTTGAAAAATATACTCTGCTCAAAAGAGACGATTCAGTTTAAACAG 1203  
Db |||||  
1313 CAGTTGATGAATTAAGAAAGTATTAATTCACGCAAGACGAGCGCTTAGTTTAAAGCAG 1372  
Qy |||||  
1204 GGATTACAGGTCCTCGCAGGTTAGTGGTCTAGTAAATATACAGACTTCGACGACGTAG 1263  
Db |||||  
1373 GAATCACTGGTTTGTGCAAAATATCTGGTGAATAATATTAATCTGATTTTGAATATCG 1432  
Qy |||||  
1264 TTCGGTTGGACTTAGCATACATGATTAATTTGGACTATCTGGTACAGATTAATAATTTAT 1323  
Db |||||  
1433 TAAAGTTAGATGTTTCAATATATCAATGAATGCTATTTTGGTACAGATTAAGATATTC 1492  
Qy |||||  
1324 TAAAGACAGTGAAGTTGTTATGTTGAGAGAGGGAAGTAAAGTAT----- 1373  
Db |||||  
1493 TCCTAACATTAAGGTAGTCTTACTCTGGACAGAGCTAAGTAAAGTTAGGTTTGAAG 1552  
Qy |||||  
1374 -----ATGAAAGTTGTTTGGTTCGTTTCAGGGGACATTTGACTCACCTTGATTTT 1426  
Db |||||  
1553 GAATATAATGAATTTGCTGGTTGTTCAAGTGGTGGTCACTAGCACACTTGAACCT 1612  
Qy |||||  
1427 GTTAAACCGTTTGGAAAGAGAAAGAGTGTGTTGGGTAAACATTTGATAAAGAGATGC 1486  
Db |||||  
1613 TTTGAAATCCATTTGGGAAAAGAGATAGGTTTGGGTAACTTTGATAAAGAGATGC 1672  
Qy |||||  
1487 AAGAAGTCTTTTGAAGATGAAAAATGATCCATGTTACTTTCCCAAAATCGCAATCT 1546  
Db |||||  
1673 TAGGAGTATTTAAGAGAGAGATTGTATATCATTTCTTTCCAAACCCGTAATGT 1732  
Qy |||||  
1547 CATTAATTTAGTGAATAATCTTTCTTAGCTTTTCAAAATTTTACGTGATGAGAAACGAGA 1606  
Db |||||  
1733 CAAAAACTGGTAAAAAATACTATCTAGCTTTTAAAGTCTCTAGAAAGAAAGACGAGA 1792  
Qy |||||  
1607 TGTATTAATTTATCTGTTGGCGGCTGCTGCTCCCTCTTTTACATCGGAAACATTT 1666  
Db |||||  
1793 TGTATCATATCATCTGGTGGCGGTAGCAGTACCATTCTTTTATATTTGTTAGTTATTT 1852  
Qy |||||  
1667 TGGAGCAAGACGATTTATATTTGAAGTATTTGATCGAGTTAATAAATCACTAACTGG 1726  
Db |||||  
1853 TGGTTGTAAAGACCAATTTATATAGAGTTTTCGACAGGATAGATAACCACTTTTGAAG 1912  
Qy |||||  
1727 AAAAAGTATTTATCCCGTAAACAGATATTTTATTTGTTTCAAGTGGGAAGAAATGAAGAAGT 1786  
Db |||||  
1913 AAAATTAGTGTATCTCTGTAACAGATAAATTTATTTGTTTCAAGTGGGAAGAAATGAAGAAGT 1972  
Qy |||||  
1787 ATATCTTAATCTATTAATCTTTGGGAGATTTTATTTTAAATGATTTTGTAAACGATAGGA 1846  
Db |||||  
1973 TTATCTTAAGCAATTAATTTAGGAGGAATTTTAAATGATTTTGTACAGTGGGAGCA 2032  
Qy |||||  
1847 CATGAACAAACAGTTTATTCGATTTGATAAAGAGATTTGATTTTATTTGAAAAAATGGAAGT 1906

2033 CATGAACAGCAGTTCAACCGTCTTATTAAGAAGTTGTAGATTAAGAAGGACAGGTGCT 2092  
Qy |||||  
1907 ATAACCGAGCAATATTTATTTCAACAGGATATTTCTGACTATTTCCAGATATTTGCAAG 1966  
Db |||||  
2093 ATTGATCAAGAAGTGTTCATTTCAACGGGTTACTCAGACTTTGAACCTCAGAATTTGTCAG 2152  
Qy |||||  
1967 TATAAAAAATTTCTCAGTTTACAAAGAAATGGAACAATATATTAACAAATCAGAAGTAGTT 2026  
Db |||||  
2153 TGGTCAAAATTTCTCATATGATGATGAACTGTTTACATGAGAGAAGCTGAGATTGTT 2212  
Qy |||||  
2027 ATTTGCCACGGAGG 2040  
Db |||||  
2213 ATCACACATGGCGG 2226

RESULT 11  
US-10-804-408-167  
; Sequence 167, Application US/10804408  
; Publication No. US20040253617A1  
; GENERAL INFORMATION:  
; APPLICANT: FANRONG, KONG  
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI  
; FILE REFERENCE: 675002-2001  
; CURRENT APPLICATION NUMBER: US/10/804,408  
; PRIOR FILING DATE: 2004-03-19  
; PRIOR APPLICATION NUMBER: PCT/AU02/01281  
; PRIOR FILING DATE: 2002-09-18  
; PRIOR APPLICATION NUMBER: AU PR 7749  
; PRIOR FILING DATE: 2001-09-19  
; NUMBER OF SEQ ID NOS: 182  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 167  
; LENGTH: 2226  
; TYPE: DNA  
; ORGANISM: Streptococcus agalactiae  
US-10-804-408-167

Query Match 9.3%; Score 652.4; DB 18; Length 2226;  
Best Local Similarity 59.6%; Pred. No. 2.6e-82;  
Matches 1225; Conservative 0; Mismatches 791; Indels 38; Gaps 6;

Qy 16 TGCATATTTGATATGATAGCAGTTGCAATTTCTGCAATCTTAACAAGTCATATACCA 75  
Db 182 TAGCGATGATTCAAACAGTTGTAGTTATTTTCTGCAAGTTTGACATTAACATTAATTA 241  
Qy 76 ATGCTGATTTAAATCG---TTCTGGAATTTTATCATATGATGGTTTCATTTTTCAT 132  
Db 242 CTCCCAATTTAAAGCAATAAAGATTTATGTTGTTCTTATGATACATTTATTTGCT 301  
Qy 133 TTTTATATCTCGTATGCGAGTTGAATTTGAGTATAGAGTAATCTGATAGAGTTGAAA 192  
Db 302 TTTATCTTCTGATTTTACAGAGACTTTTGGAGTCTGGCTATCTTGAAGAGTTTAAAA 361  
Qy 193 AAACATTTAACTATAGTATTAATTTGCAATTTTCTTACGGCAGTATCATTTTGTGG 252  
Db 362 TGGTATGGAATAACAGCTTTTACTATATTTTCAATCAAGTTCAATTTTATTTTATTTA 421  
Qy 253 AGAATAATTTGCGACCTTTCAAGACGTTGGTCCGTGATTTTCAATTAATAAATTCGTTT 312  
Db 422 AAACTCTTTTACACGACGACTTTCCTTTTACTTTTATGCTATGAAATTCGATTT 481  
Qy 313 TGGTATACCTTATTAACGTAATTTTAAAGCAGTTTAAAGGATAGCTTTCTATTTTCGCAA 372  
Db 482 TATTGATCTTATTGAATTCATTTTAAAAATATTTATCGAAAAATATCTTACGCTAAGTTT 541  
Qy 373 TCTATCAAAAAAGACGATTTCTAATTTACACGCTGAACGATGGAAAAATATGCAAGTTT 432  
Db 542 CACGAGATACCAAGTTGTTTGTATACGAATTAAGGATTTCTTTATCAAAAATGACCTTTA 601  
Qy 433 TATTGGAATCACATAAACAAATTTCAAAAAATCTTGTGCAATTTGGTAGTTTGAAGTACAG 492

Db 602 GGAATAAT---ACGACCAATATATATCGCTGTCTGTATCTTGACTCTCTCTGAAAAGG 658  
Qy 493 AAATAGATAAAATTAATTAATCAATACCGCTCTATTAATTTCTGTGGAAGAAGCATATAGAGT 552  
Db 659 ATTGTTATGATTTGAAAACATAACTCGTTAAGGATAATAAACAAGATGCTCTTACTTCAG 718  
Qy 553 TTTCAACAAGGAAGTGTGCGACGAGCTTTTATAAATCTACCAAGTGAAGTTTGTAG--- 609  
Db 719 AGTTAACTGCTTAACCTGTTGATCAAGCTTTTATTAACATACCCATTGAATTAATTTGGTA 778  
Qy 610 -----ACGTAAGCAATTCGTTTCAGATTTTGAAGTTTGTAGTATTTGATGTAAGCGTTG 663  
Db 779 AATACCAATACAGATTAATTAATGACATTTGAAGCAATGCGAGTGAATGTCATGTTA 838  
Qy 664 ATATTAATTCATTCGGTTTACTCGGTTGAAAAAACAATAATCAACTGCTAGTGACC 723  
Db 839 ATGTAGAGCACTTAGCTTTGATAATATAGGAGAAAAGCGAATCCAACTTTTGAAGGAT 898  
Qy 724 ATAGCAATTTGTAATTTTCCCAAAATTTTATAAGCCTAGTCATATCATGATGAACGAC 783  
Db 899 ATAGTGTATTACATATTTCTATGAATTTCTATAAATATAGTCACCTTATAGCAAAACGAT 958  
Qy 784 TTTTCGATATACTCGGAGCGGTAGTCGGTTAAATTTATTTGTTGATAGTTTCTTATTTGT 843  
Db 959 TTTTGGATATACGGGTGCTATTATAGGTTTCTCATATGTGGCAATGTGGCAATTTTTC 1018  
Qy 844 TAGTTCCAAATTAATTCGTAGAGATGTGGACCGGCTATTTTGTCTCAGAAAACGAGTTGGAC 903  
Db 1019 TAGTTCCAAATCAGAAAAGATGTGGACCGGCTATCTTTTCTCAAAATAGAGTAGTTC 1078  
Qy 904 AGAATGGAGCAATTTTACATCTCAAGTTTCGATTCGATGATGTTGATGCTGAGGAGC 963  
Db 1079 GTAATGTTAGGATTTTGAATTTCTATAAATTCAGATCAATGCGAGTAGATGCAGAACAA 1138  
Qy 964 GCAAAAAGACTGTCTGACCCAAACACAGATGCAAGGTTGGGTATGTTTAAATGGGAA 1023  
Db 1139 TTAAGAAGATTTATAGTTCAATCAATCAATGACAGGG-----CTAATGTTTAAAGTTAGA 1193  
Qy 1024 AAACGATCCTAGAAATTAATCCAAATTTGGACATTTTCATAGCCAAAAACAAGTTTAGACGAGT 1083  
Db 1194 AGATGATCCTAGAAATTAATTAATAGGAAATTTATTCG-AAAAACAAGCATAGATGAGT 1252  
Qy 1084 TACCAAGTTTATAATGTTTAAATTTGGGATATAGAGTCTAGTTGGTACAGTCCACCTA 1143  
Db 1253 TGCTCAATTTCTATAATGTTTAAAGGTGATATGAGTTTAGTGAACAAGCCCTCCCA 1312  
Qy 1144 CAGTTGATGAATTTGAAAAATATACCTCTGTCGTAAGAGAGAGAGTGTAGTTTAAACAG 1203  
Db 1313 CAGTTGATGAATATGAAAAGTATAATTCACGCGAAGCGAGCGCTTAGTTTAAAGCCAG 1372  
Qy 1204 GGATTAAGGTTCTGCGCAGGTTAGTGTGCTAGTAATATACAGACTTCGACGACGATG 1263  
Db 1373 GAATCAGCTGTTTGGGCAATATCTGGTAGAATAATATCACTGATTTGATGAATTCG 1432  
Qy 1264 TTCGGTTGGACTTAGCATACATTTGATAATTTGGACTATCTGCTCAGATATTAATAATTTAT 1323  
Db 1433 TAAAGTTAGATGTTCAATATATCAATGAATGCTCTATTTTGGTCAGATATTAAGATTATC 1492  
Qy 1324 TAAAGACAGTGAAGTTGTTGTTGAGAGAGGGAAGTAAGTAAAGTAT----- 1373  
Db 1493 TCCTAACACATAAGGTTAGTCTTACTTGGGACAGGTTGCTAAGTAAAGGTTAAGGTTTGAAG 1552  
Qy 1374 -----ATGAAAGTTTGTGTCGTTCTTCAGGGGACATTTGACTCACTCTATTT 1426  
Db 1553 GAATATAATGAATAATTTGCTGTTGTTCAAGTGGTGTCTATCTAGCACACTTTGAACCT 1612  
Qy 1427 GTTAAACCGTTTGGAAAGGAAGAACGTTTGGGTAAATTTGATTAAGAGAGATGC 1486  
Db 1613 TTTGAAACCATTTGGGAAAGAGATAGGTTTGGGTAACTTTGATTAAGAGATGC 1672  
Qy 1487 AAGAGCTTTTGAAGAGATAAATAATGATTCATGTTTCAACAAATTCGCAATCT 1546  
Db 1673 TAGGAGTATTTCTAAGAGAAGAGATTGTATATCATTTGCTTCTTCCAAACAAACCGTAATGT 1732

Qy 1547 CATTAAATTTAGTGAAAAATACCTTTCTTAGCTTTCAAAATTTTACGATGAGAAACCAGA 1606  
Db 1733 CAAAAACTTGGTAAAAAATACCTATTCTTAGCTTTTAAAGTCTCTTGAAGAAGAACCCAGA 1792  
Qy 1607 TGTATTAATTTCACTGCTGCGGCGTGTCTGCTCCCTTCTTTTACATCGGAAACTATT 1666  
Db 1793 TGTATCATATCATCTGCTGCGCTGTAGCAGTACCAATTTCTTTTATATTTGGTAAGTTAT 1852  
Qy 1667 TGGCAAGACAGATTTATATTTGAAGTATTTGATCGAGTTAATAATCTACATTAACCTGG 1726  
Db 1853 TGGTTGTAAGACCGTTTATATAGAGTTTTCGACAGGATAGATAAACCAACTTTGACAGG 1912  
Qy 1727 AAACTAGTTTATCCCGTAAACAGATATTTTATTTGTTTCACTGCGGAGAAATGAAGAAGT 1786  
Db 1913 AAAATTAGTGTATCTGTAAACAGATAAATTTATTTGTTCACTGCGGAGAAATGAAAAAGT 1972  
Qy 1787 ATATCTTAATCTATTAACTTTGGGAGTATTTTAAATGATTTTGTAAACAGTAGAACT 1846  
Db 1973 TTATCTTAAGCAATTAATTTTAGGAGGAATTTTAAATGATTTTGTCAAGTGGGACA 2032  
Qy 1847 CATGAACAAACAGTTTAACTCGATTGATAAAAGAGATTTGATTTTTCGAAAAAATCGAAGT 1906  
Db 2033 CATGAACAGAGTTCAACCGTCTTATTAAAGAGTTGATAGATTTAAAGGACAGGTGCT 2092  
Qy 1907 ATAAACCGACGAATATTTTATTTCAACAGGATATTTCTGACTATATTTCCAGAAATATTGCAAG 1966  
Db 2093 ATTGATCAAGAAGTGTTCATTTCAACCGGTTACTCAGACTTTGAACCTCAGAAATTTGTCAG 2152  
Qy 1967 TATAAAAATTTCTCAGTTTACAAAGAAATGGAACAATATATTTAAACAATTCAGAAAGTAGTT 2026  
Db 2153 TGGTCAAAATTTCTCATATGATGATGAATGAATCTTTACATGAAGAAGCTGAGATTGTT 2212  
Qy 2027 ATTTGCCACGAGG 2040  
Db 2213 ATCACATGCGG 2226

## RESULT 12

US-10-804-408-162  
; Sequence 162, Application US/10804408  
; Publication No. US20040253617A1  
; GENERAL INFORMATION:  
; APPLICANT: FANRONG, KONG  
; APPLICANT: GILBERT, GWENDOLYN  
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI  
; FILE REFERENCE: 675002-2001  
; CURRENT APPLICATION NUMBER: US/10/804,408  
; PRIOR FILING DATE: 2004-03-19  
; PRIOR APPLICATION NUMBER: PCT/AU02/01281  
; PRIOR FILING DATE: 2002-09-18  
; PRIOR APPLICATION NUMBER: AU PR 7749  
; PRIOR FILING DATE: 2001-09-19  
; NUMBER OF SEQ ID NOS: 182  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 162  
; LENGTH: 2217  
; TYPE: DNA  
; ORGANISM: Streptococcus agalactiae  
US-10-804-408-162

Query Match 9.3%; Score 650.8; DB 18; Length 2217;  
Best Local Similarity 59.6%; Pred. No. 4.3e-82;  
Matches 1224; Conservative 0; Mismatches 792; Indels 38; Gaps 6;  
Qy 16 TGGCATTATTGTATGATAGTAGCAGTTGCAATTTCTGCAATCTTAAACAGTCATATACCAA 75  
Db 173 TAGCGATGATTCAACACMGTTGCTAGTTTATTTTCTGCAAGTTTGACATTAACATTAAATTA 232  
Qy 76 ATGCTGATTTAAATCG---TTCTGGAATTTTATCATATGATGTTTCAATTTATTTTGCAT 132  
Db 233 CTCCTCAATTTTAAAGCAATAAAGATTATTTGTTTGTCTTATGATACATATATTTGTTT 292

Qy	133	TTTTTATATCTCGPATGCCAGTTGAAATTTGAGTATAGAGGTAATCTGATAGAGTTTGA	192
Db	293	TTTTATCTTTCTGATTTTTTACAGAGACTTTTTGGAGTCGTGCTATCTTGAAGAGTTTAA	352
Qy	193	AAACATTTAACTATAGTATAATATTTGCCAATTTTCTTACGGCAGTATCATTTTTTGTGG	252
Db	353	TGGTATTGAAATACAGCTTTTACATATATTTTCATATCAAGTTCATTTATTTTATTTT	412
Qy	253	AGAATAATTTTCGCACTTTCAAGACGTGGTGCCTGTATTTTACATTAATAAACTTCGTTT	312
Db	413	AAAACCTTTTACAAACGACGACTTTTCCTTTTACCTTTTATTTGCTATCGATTCGATTT	472
Qy	313	TGGPATACCTATTTPAACGTAAATTAATAAGCAGTTTAAAGTAGAGTCTTCTATTTTCG	372
Db	473	TATTTATCTATTGAAATTCATTTTAAATATTTATCGAAAATATCTTACGCTAAGTTT	532
Qy	373	TCATCAAAAAAAGACAGATTTCTAATTTACACGGCTGAACGATGGGAAAAATATGCAAGT	432
Db	533	CACGAGATACCAAGTTGTTTGTGATACGAAATAGGATTCCTTTATCAAAAATGACCTTT	592
Qy	433	TATTTGAATCACATAAAACAAATTCAAAAAATCTGTTCATTTGGTAGTTTATAGGTACAG	492
Db	593	GGAAATAAT--ACGACCAATTAATATATCGCTGCTGCTATCTTGGACTCTCTGAAAAGG	649
Qy	493	AAATAGATAAAATTTAATTTATCATTCGCTCTATTTATTTCTGTGGAAAGAGCTATAGAGT	552
Db	650	ATTGTTATGATTTGAAACATAAATCGTTTAAAGGATAATAAAACAAAGATGCTTTACTT	709
Qy	553	TTTCAACAAAGGGAAGTCGTGACACAGCTCTTTTATAAATCTACCAAGTGAGTTTTTAG--	609
Db	710	AGTTAACTCGCTTAACCTGTTGATCAAGCTTTTATTAACATACACCAATTTGAATTTAT	769
Qy	610	-----ACGTTAAAGCAATTCGTTTTCAGATTTTGAGTTTGTAGGTATTTGATGTAAGCG	663
Db	770	AATACCAAAATACAAGATTAATTAATGACATTTGAAGCAATGGGAGTGATTTGTCAAITG	829
Qy	664	ATATTAATTCATTCGGTTTTTACTCGTTGAAAAACAAAAATCCAACTGCTAGGTGACC	723
Db	830	ATGTAGAGGCATTTAGCTTTTGATATAATAGGGAAGACGNAITCCAACTTTTGAAGAT	889
Qy	724	ATAGCATGTAACTTTTTTCCACAAATTTTATAAGCCTAGTGCATPATCATGATGAAACG	783
Db	890	ATAGTGTATTACATAATTTCTATGAAATTTCTATAAATATAGTCACTTTATAGCAAAAC	949
Qy	784	TTTTGGATATACTCGGAGCGGTAGTCGGGTTAATTTATTTGCGTATAGTTTCTATTTTGT	843
Db	950	TTTTGGATATCACGGGTGCTATTATAGGTTTGTCTCATATGCGGCAATTTGTGCGCAAT	1009
Qy	844	TAGTTCCAAATTTATTCGTAGAGATGGTGGACCGGCTATTTTGTCTCAGAAAACGAGT	903
Db	1010	TAGTTCCGCAATCAGAAAAGATGGTGGACCGGCTATCTTTCTCAAAATAGAGTAGGTC	1069
Qy	904	AGAATGGACGCAATTTATCATCTTACAAAGTTTCGATCGATGTATGTTGATGCTGAGG	963
Db	1070	GTAATGSGTAGGATTTTTTAGATTCTATAAAATTCAGATCAATCAATGCGAGTAGATG	1129
Qy	964	GCAAAAAGACTTCCTCAGCAAAACAGATGCAAGGTGGGTATGTTTAAATGGGAA	1023
Db	1130	TTAAGAAGATTTTATAGTTTACCAATCAAAATGACGGG-----CTAATGTTTAA	1184
Qy	1024	AAACGATCCTAGAATTAATCTCAAATGGACATTTTCATACGCAAAAACAAAGTTT	1083
Db	1185	AGATGATCCTAGAAATTAATAAATAGGNAATTTTATTCG--AAAAACAAAGCATA	1243
Qy	1084	TACCACAGTTTTTATAATGTTTTTAATTTGGCGATATAGTCTAGTTGGTACACGCTC	1143
Db	1244	TGCCTCAATTCATAATGTTTTTAAAGGTGATATGAGTTTAGTAGGAAACACGCCCT	1303
Qy	1144	CAGTTGATGAATTTGAAAAATATATCTCCTGGTCAAAAGACGATTTGAGTTTTTAA	1203
Db	1304	CAGTTGATGAATATGAAAAGATTAATTTCAACGCAAGACGACGCTTAGTTTAA	1363
Qy	1204	GGATTAACAGGTCCTCGGCAGGTTAGTGGTCGTAGTAAATATACAGACTTCGACAC	1263

RESULT 13

RESULTS 13  
US-10-804-408-173

US-10-804-408-173  
; Sequence 173, Application US/10804408

US 20040253617A1

; PUBLICATION NO. US20  
; GENERAL INFORMATION:

; APPLICANT: FANRONG, KONG

APPLICANT: FANKONG, KONG  
APPLICANT: GILBERT, GWENDOLYN

APPLICANT: GILBERT, GWENDOLYN  
TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI

FILE REFERENCE: 675002-2001

; FILE REFERENCE: 075002 2001  
; CURRENT APPLICATION NUMBER: US/10/804,408

;; CURRENT AFFILIATION NUMBER: 03/10/004,408  
;; CURRENT FILING DATE: 2004-03-19

;; CURRENT FILING DATE: 2004-03-15  
;; PRIOR APPLICATION NUMBER: PCT/AU02/01281

; PRIOR AFFIDAVIT NUMBER: FCI  
 ; PRIOR FILING DATE: 2002-09-18

;; PRIOR APPLICATION NUMBER: AU PR 7749  
;; PRIOR FILING DATE: 2001-09-19  
;; NUMBER OF SEQ ID NOS: 182  
;; SOFTWARE: PatentIn version 3.2  
;; SEQ ID NO 173  
;; LENGTH: 2226  
;; TYPE: DNA

;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: Consensus sequence

US-10-804-408-173

Query Match 9.3%; Score 649.2; DB 18; Length 2226;

Best Local Similarity 59.5%; Pred. No. 7.3e-82;

Matches 1223; Conservative 0; Mismatches 793; Indels 38; Gaps 6;

Qy	16	TGGCAATATTGATGATAGCAGTTGCAATTTCTGCAATCTTAACAGTCAATATACCAA	75
Db	182	TAGCGATGTTCAACAGTTGTGTTTATTTTCTGCAAGTTGACATTAACATTAATTA	241
Qy	76	ATGCTGATTTAAATCG---TTCTGGAATTTTATCATATGATGGTTCATTTATTTGCA	132
Db	242	CTCCCAATTTAAAGCAATAAGATTTATTTGTTGTTCTATTGATACATTTATTTGCT	301
Qy	133	TTTTTATATCTGATGCGAGTTGAATTTGATGATAGAGTAATCTGATAGATTTGAAA	192
Db	302	TTTATCTTTCTGATTTTACAGAGACTTTTGGAGTCGTGGCTATCTTGAAGAGTTTAAA	361
Qy	193	AAACATTTAACTATAGTATATATTTGCAATTTTCTTACGGCAGTATCATTTTGTGG	252
Db	362	TGGTATTGAAATACAGCTTTTACTATATTTTTCATATCAAGTTTCATTTTATTTT	421
Qy	253	AGATTAATTTCCGCACTTTCAAGAGCTGTGGCGGTATTTTTCACATTAATAAATCTG	312
Db	422	AAACTCTTTTACACGACAGCTTTCTCTTTTACTTTTATTTGCTATGATTCGATTT	481
Qy	313	TGGTATACCTATTAACGTAATTAATAGCAGTTTAAGGATGCTTCTATTTTCGACAA	372
Db	482	TATTATATCTATGAAATTCATTTTAAATATTTATCGAATAATTTCTTACGCTAAG	541
Qy	373	TCATCAAAAAAGACGATTTCTAATTACACGGCTGAACGATGGGAATATCAAGTTT	432
Db	542	CACGAGATACCAAGTTGTTTGTGATAACGAATAAGGATTTCTTATCAAAAAATG	601
Qy	433	TATTTGAATCACATAACAAATTCAAAAAATCTGTTGCAATGTTAGTATTTAGTAC	492
Db	602	GGATTAAT---ACGACCTAATTAATATCGCTGCTGATCTTGACCTCTGAAAGG	658
Qy	493	AAATAGATAAAATTAATTTATCATTCACGCTCTATTTATTTCTGGAAGAAGCTAT	552
Db	659	ATTGTTATGATTTGAAACATAACTCGTTAAGGATAATAACAAAGATGCTCTTACT	718
Qy	553	TTTCAACAGGAAGTGTGACACGCTTTTATAAATCTACCAAGTGAGTTTATAG---	609
Db	719	AGTTAACCTGCTTAACCTGTTGATCAAGCTTTTATTAACATACCCCATTTATTTG	778
Qy	610	-----ACGTAAGCAATTCGTTTCAGATTTTGTAGTTGTTAGTATTTGATGTA	663
Db	779	AATACCAATACAGATTAATTAATGACATGGAAGCAATGGAGTGATTTGATGTTA	838
Qy	664	ATATTAATTCATTCGTTTAACTGCGTTTGAATAACAAAAATCAACTGCTAGTGACC	723
Db	839	ATGATAGGCACTTAGCTTTGATAATATAGGAGAAAGCAATCCAACTTTTGAAGAT	898
Qy	724	ATAGCATTTGAATCTTTTCCAAATTTTATAAGCCCTAGTCAATCATGATGAAACG	783
Db	899	ATAGTGTATTACATATTTCTATGAAATTTCTATAAATATAGTCACCTTATAGCA	958
Qy	784	TTTTGGATATCTCGGAGGCTAGTCGGGTTAAATTTATTTGTTGATAGTTCTATT	843
Db	959	TTTTGGATATCACGGGCTGCTATTATAGGTTTGTCTCATATGTTGGCAATTTTTC	1018
Qy	844	TAGTTCCAAATTTATTCGTAGATGTTGGACCGGCTATTTTGTCTCAGAAACGAG	903

Db	1019	TAGTTCCGCAAAATCAGAAAAAGATGGTGACCGGCTATCTTTTCTCAAAATAGAGTAG	1078
Qy	904	AGATGAGCGCATATTTTACATTTCTACAAGTTTTCGATCGATGATGTTGCTGAGGAG	963
Db	1079	GTAAATGTTAGGATTTTATGATTTCTATAAATTCAGATCAATCGGAGTAGATGAG	1138
Qy	964	GCAAAAAAGACTTGTCTCAGCCAAAAACAGATGCAAGGGTGGGTATGTTTTAAAT	1023
Db	1139	TTAAGAAAGATTTTATAGTTTCACAATCAATGACAGGG-----CTAATGTTAAG	1193
Qy	1024	AAACGATCTAGATTAATCTCAATTTGGACATTTCTATCGCAAAACAGTTTACAG	1083
Db	1194	CGATGATCTAGAAATTAATAAATAGGAAATTTATTCG-AAAAACAAGCATAGAT	1252
Qy	1084	TACCACAGTTTATAATTTTAAATTTGCGCATATGAGTCTAGTTGGTACACGTCAC	1143
Db	1253	TGCTCTCAATTTCTATAATTTTAAAGGTGATATGAGTTAGTAGGAACAACGCTCC	1312
Qy	1144	CAGTTGATGAATTTGAAAAATATATCTCTGGTCAAAAGAGAGAGATTTGAGTTT	1203
Db	1313	CAGTTGATGAATATGAAAAAGTATAATTCACGCAAGAGCGAGCTTAGTTTAA	1372
Qy	1204	GGATTACAGGCTCTGCGCAGGTTAGTGGTCTGATGATATATCAAGCTTCACGAC	1263
Db	1373	GAATCACTGGTGTGCGCAATATCTGGTAGAAATAATATTACTGATTTTGTGAAAT	1432
Qy	1264	TTCCGTTTGGACTTAGCATACATTCATTAATTTGGACTATCTGGTCAAGATTA	1323
Db	1433	TAAAGTTAGATGTTCAATATATCAATGATGCTCTATTTGGTCAAGATTAAGAT	1492
Qy	1324	TAAACAGCTGAAAGTTGTTTGTGAGAGAGGAAAGTAAAGTAAAGTAT-----	1373
Db	1493	TCCTAAACACTAAAGGTAGTCTTACTCGGACAGGAGCTAAGTAAAGGTTAAAG	1552
Qy	1374	-----ATGAAAGTTTGTGGTGGTCTTTCTAGGGGCAATTTGACTCACTGTAT	1426
Db	1553	GAATATAATGAAATTTGCTGGTGGTCAAGTGGTGGTCTAGCAGCACTTTGAAC	1612
Qy	1427	GTTAAAAACGTTTGGGAAGGAAGAAACGTTTGGTGAACATTTGATAAAGAGGAT	1486
Db	1613	TTTGAACCCCATTTGGGAAAGAAAGATAGGTTTGGTGAACCTTTGATAAAGAG	1672
Qy	1487	AAGAAGCTTTTGAAGAAATGAAAAAATGATTCATTTCTTCCAAACAAATCGAAT	1546
Db	1673	TAGGAGTATTTCTAGAGAGAGAGATTTGATATCATTTCTTCTTCCAAACAA	1732
Qy	1547	CATTAATTTAGTGAATAATCTTTCTAGCTTTTCAAAATTTTACGATGAGAAAC	1606
Db	1733	CAAAAACTTGGTAAAAAATACTATTCTAGCTTTTAAAGTCTTTAGAAAAAGAA	1792
Qy	1607	TGTTATTTATCTCGTGGCGGCTTGTGTCCTTCTTCTTCTTCTTCTTCTTCTT	1666
Db	1793	TGTTATCATATCATCTCGTGGCGCTGTAGCAGTACCATTTCTTTATTTTAT	1852
Qy	1667	TGAGACAAAGACGATTTATATTGAAGTATTTGATCGAGTTTAAATCTAATTA	1726
Db	1853	TGGTTGTAAGACCGTTTATATAGAGTTTTCGACAGGATAGATAAACAACCTT	1912
Qy	1727	AAAACTAGTTTATCCCGTAAACAGATATTTTATTTGTTTCAAGTGGGAAGAA	1786
Db	1913	AAAAATAGTGTATCTCTGTAACAGATAAATTTATTTGTTTCAAGTGGGAAGAA	1972
Qy	1787	ATATCTTAATCTATTAACTTGGGAGTATTTTATTTTATTTTATTTTATTTT	1846
Db	1973	TTATCTTAAGCAATTAATTTAGGAGAAATTTTAAATGATTTTGTCTACAGT	2032
Qy	1847	CATCAACAAACAGTTTAAATCGATTCATATAAAGAGATTTGATTTTATTTT	1906
Db	2033	CATGACACAGGTTTCAACCGCTTTTAAAGAGTTTGNATGATTTAAAGGAGCA	2092
Qy	1907	ATAACCGACGAAATTTATTTTCAACAGGATTTCTGCTATATTTCCAGATA	1966

Db 2093 ATTGATCAAGAGTGTTCATTCAAACGGGTACTCAGACCTTGAAACCTCAGAAATGTCAG 2152  
Qy 1967 TATATAAAATTTCTCAGTTACAAGAAATGGAACAATATATATTAACAAATCAGAAGTAGTT 2026  
Db 2153 TGGTCAAAATTTCTCATATGATGATGAATGAATCTTTACATGAAGAAGCTGAGATTGTT 2212  
Qy 2027 ATTTGCCACGAGG 2040  
Db 2213 ATCACATGCGG 2226

RESULT 14  
US-10-804-408-171  
; Sequence 171, Application US/10804408  
; Publication No. US2004025361/7A1  
; GENERAL INFORMATION:  
; APPLICANT: FANRONG, KONG  
; APPLICANT: GILBERT, GWENDOLYN  
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI  
; FILE REFERENCE: 675002-2001  
; CURRENT APPLICATION NUMBER: US/10/804,408  
; CURRENT FILING DATE: 2004-03-19  
; PRIOR APPLICATION NUMBER: PCT/AU02/01281  
; PRIOR FILING DATE: 2002-09-18  
; PRIOR APPLICATION NUMBER: AU PR 7749  
; PRIOR FILING DATE: 2001-09-19  
; NUMBER OF SEQ ID NOS: 182  
; SOFTWARE: Patent in version 3.2  
; SEQ ID NO 171  
; LENGTH: 2226  
; TYPE: DNA  
; ORGANISM: Streptococcus agalactiae  
; US-10-804-408-171

Query Match 9.3%; Score 648.8; DB 18; Length 2226;  
Best Local Similarity 59.5%; Pred. No. 8.3e-82;  
Matches 1222; Conservative 1; Mismatches 793; Indels 38; Gaps 6;

Qy 16 TGGCATTAATTGATGATAGCAGTTCGAATTTCTGCAATCTTAACAAGTCATATACCAA 75  
Db 182 TAGCGATGATTCAAACMGTTGTGGTTTATTTCTGCAAGTTTGACATTAACATTAATTA 241  
Qy 76 ATGCTGATTTAAATCG---TTCTGGAATTTTATCAATAGATGTTTCATATTTTGCAT 132  
Db 242 CTCCCAATTTAAAGCAATAAAGATTTATGTTGTTCTTATTGATACATATATTGTCT 301  
Qy 133 TTTTATATCTCGATGCGAGTTCAATTTGAGTATAGAGGTAATCTCATAGAGTTTGAAA 192  
Db 302 TTATCTTTCTGATTTTACAGACATTTTGGAGTCGTGGCTATCTTGAAGAGTTTAAAA 361  
Qy 193 AAACATTTAACTATAGTATATATTTTGCATTTTCTTACGGCAGTATCATTTTGTGG 252  
Db 362 TGGTATTGAAATACAGCTTTTACTATATTTTCAATCAAGTTTCATTTTATTTTATTTTA 421  
Qy 253 AGATAATTTTCGACTTTCAAGCGTGGTCCGCTGATTTTCAATTAATCAACTTCGTTT 312  
Db 422 AAAAATCTTTTACACGACAGCTTTTCTTTTACTTTTATTTGCTATGAAATTCGATTT 481  
Qy 313 TGGTATACCTATTTAACTGATTAATTAAGCAGTTTAAAGGATAGCTTTCTATTTTCGACAA 372  
Db 482 TATTATATCTATGAAATTCATTTTAAATATTAATCGAAATATCTTACGCTAAGTTT 541  
Qy 373 TCTATCAAAAAGACGATTTCTAATTAACCGCTGAACGATGGGAAAAATATGCAAGTTT 432  
Db 542 CACGAGATACCAAAGTTGTTTGTGATAACGAATAGGATCTTTTATCAAAAATGACCTTTA 601  
Qy 433 TATTGATCACATAAACAATTCAAAAATCTTTGTTGCAATGCTAGTTTATTAGTACAG 492  
Db 602 GGAATAAAT---ACGACCATTAATATATCGCTGTCTGTATCTTGGATTCCTCTGAAAAGG 658  
Qy 493 AAATAGATAAATTTAATTTATCATTTACCGCTCTATTTTCTGTGGAAGAAGCTATAGT 552  
Db 659 ATTGTTATGATTGAAACATCACTCGTTAAGGATAATAAACAAGATGCTCTTACTTCAG 718

Qy 553 TTTCAACAAGGAGTGGTCGACCAGCTCTTTATATAATCTACCAAGTGAGTTTTTAG--- 609  
Db 719 AGTTAACTTCGCTTAACCTGTTGATCAAGCTTTTATTAACATACCCATTGAATTAATTTGGTA 778  
Qy 610 -----ACGTAAAGCAATTCGTTTTCAGATTTTGGTGTAGTATTTGATGTAAGCGTTG 663  
Db 779 AATACCAATACAGATATTTAATGACATTTGAAGCAATGGAGTGATTTGTCATGTTA 838  
Qy 664 ATATTAAATTCATTTCGGTTTTTACTGCGTTTGAAAAACAAAAAATCCAACTGCTAGGTGACC 723  
Db 839 ATGTAGAGGCACCTTAGCTTTGATAATATAGGAGAAAAAGCGAATCCAAAACCTTTTGAAGGAT 898  
Qy 724 ATAGCATTTGAATCTTTTCCACAAATTTTATAAGCCTAGTCTATATCATGATGATAAAGAC 793  
Db 899 ATAGTGTATTACATATCTTATGAATTTCTATAAATATAGTCACTTTATAGCAAAACGAT 958  
Qy 784 TTTTGGATATACTCGGAGCGGTAGTCGGGTAAATTTTGTGTTATTTGTGTTATTTTCTATTGTT 843  
Db 959 TTTTGGATATCAGGGTGTCTATTATAGGTTTGTCTATATGTGGCATTTGGCAATTTTTC 1018  
Qy 844 TAGTTCCAAATTTCTGATAGATGGTGGACCGGCTATTTTGTCTAGAAACGAGTTGGAC 903  
Db 1019 TAGTTCCACAAATCAGAAAAAGATGGTGGACCGGCTATCTTTCTCAAAATAGAGTAGGTCT 1078  
Qy 904 AGAATGGACGCATATTTTACATTTCTCAAGTTTCGATCGATGATGTTGATGCTGAGAGC 963  
Db 1079 GTAATGGTAGGATTTTATAGTTCTATAAATTCAGATCAATGCGAGTAGATGCAAAACAA 1138  
Qy 964 GCAAAAAGACCTGCTCAGCCAAACCCAGATGCAAGGTTGGGTATGTTTAAAATGGAA 1023  
Db 1139 TTAAGAAAGATTTATTAGTTTCAATCAATGACAGGG-----CTAATGTTTAAGTTAGA 1193  
Qy 1024 AAACGATCCTAGAAATTTACTCCAATTTGGACATTTTATACGCAAAAAACAAAGTTTAGACAGT 1083  
Db 1194 CGATGATCCTAGAAATTTACTAAAAATAGAAAAATTTATTTCG-AAAAACAAGCATAGATGAT 1252  
Qy 1084 TACCAAGTTTTATTAATGTTTTTAATTTGGCGATATGATGCTAGTTGGTACACGTCACCTTA 1143  
Db 1253 TGCCTCAATCTTATATATGTTTTAAAGGTTGATATGATGTTTAGTAGAAACACGCGCTCCCA 1312  
Qy 1144 CAGTTCATGAATTTGAAAAATATACCTCCTGGTCAAAAGAGACGATTTGAGTTTAAAACCGAG 1203  
Db 1313 CAGTTGATGAATATGAAAAAGATATAATCAACGCAAGAGCGACGCTTAGTTTAAAGCCAG 1372  
Qy 1204 GGAATTAACAGTCTCTCGCAGGTTAGTGGTCGTAGTAATATCAAGACTTTCCAGACGCTAG 1263  
Db 1373 GAATCACTGGTTTGGCAAAATATCTGGTAGAAATATATTAATCTGATTTTGTGATAATCG 1432  
Qy 1264 TTCGTTGGACTTAGCATACATTTGAATAATTTGGACTATCTGGTCAGATATTTAAAATTTTAT 1323  
Db 1433 TAAAGTTAGATGTTTCAATATATCAATGAAATGGTCTATTTTGGTCAGATATTTAAGATTTTC 1492  
Qy 1324 TAAAGACAGTGAAGTTGTTATCTTCAGAGAGGGAAGTAGTAAGTAAAGTAT----- 1373  
Db 1493 TCCTAACACTAAAGGTAGTTTACTCGGACAGAGGCTAAGTAAAGGTAGGTTTGAAG 1552  
Qy 1374 -----ATGAAAGTTTGTGTTGGTCCGTTCTTCAGGGGAGACATTTGACTCAGTTGATTTT 1426  
Db 1553 GAATATAATGAATAATTTGTTCTGGTTCAAGTGGTGGTCTATAGCACACTTTGAACCT 1612  
Qy 1427 GTTAAAAACGTTTTTGAAGGAAGAAACGTTTTTGGGTAAACATTTTGAATAAGAGGATGC 1486  
Db 1613 TTTGAAACCCATTTGGGAAAAAGAGATAGGTTTTTGGGTAAACCTTTTGTATAAGAGATGC 1672  
Qy 1487 AGAAGCTCTTTGAAGAAATCAAAAATGATCCATGTTACTTTTCCCAACAAATCGCAATCT 1546  
Db 1673 TAGGAGTATTTCTAAGAGAGAGATTGTTATATCAATGCTTTCTTCCCAACAAACCGTAAATGT 1732  
Qy 1547 CATTAATTTAGTGAAAAAATACTTTTCTAGCTTTTCAAAATTTTACGTGATGAGAAACAGGA 1606  
Db 1733 CAAAAATTTGGTAAAAAATACTATTCTAGCTTTTAAAGTCTTTAGAAAAAGAAAGACAGGA 1792

Qy	1507	TGTTATTATTTTCATCTGGTGGCGCGTTCGTGTCCTCTCTTTTACATCGGAAACTATT	1666
Db	1793	TGTTTATCATATCATCTGGTGGCGCTGTAGCAGTACCATTCTTTTATATTGGTAACTTATT	1852
Qy	1667	TGGAGCAAGACGAGTTTATATTGAAGATATTGTATCGAGTTTAATAAACTCATATTAACTGG	1726
Db	1853	TGGTTGTAAAGCCGTTTACATAGAGTTTTCGACAGGATGGATAAACCACTTTGACAGG	1912
Qy	1727	AAAACTAGTTTATCCCGTAAACAGATATTTTATTTGTTTCAGTGGGAAGAAATGAAAGAGGT	1786
Db	1913	AAAATTAGTGTATCCTGTAAACAGATAAATTTTATTTGTTTCAGTGGGAAGAAATGAAAAAGT	1972
Qy	1787	ATATCCTTAATCTATTAATTAATTTGGGGAGTATTTTTTTTAAATGATTTTGTAAACAGTAGGAAC	1846
Db	1973	TTATTCCTTAAGGCAATTAATTTTAAAGGAAATTTTTTAAATGATTTTGTACAGTGGGACA	2032
Qy	1847	CATGAACAACAGTTTAAATCGATTGATAAAAGAGATTGATTTTATTTGAAAAAATAATGGAAGT	1906
Db	2033	CATGAACAGCAGTTTCAACCGTCTTTTAAAGAAAGTTGATAGATTAAAGGACACAGGTGCT	2092
Qy	1907	ATAACGACGAAATATTTATTTCAAACAGGATATTTCTGACTATATTCAGAAATATTGCAAG	1966
Db	2093	ATTGATCAAGAAGTGTTCATTCAAACGGGTACTCAGACTTTGAAACCTCTGAAATTTGTCAG	2152
Qy	1967	TATAAAAAATTTCTCAGTTACAAAGAAATGGAACAAATATATTAACAATCAGAAAGTAGTT	2026
Db	2153	TGGTCAAAATTTCTCTCATATGATATGAATCTCTTACATGAAGAAGCTGAGATTGTT	2212
Qy	2027	ATTGGCCACGGAGG	2040
Db	2213	ATCACATATGGCGG	2226

## RESULT 15

```

US-10-804-408-172
; Sequence 172, Application US/10804408
; Publication No. US20040253617A1
; GENERAL INFORMATION:
; APPLICANT: FANRONG, KONG
; APPLICANT: GILBERT, GWENDOLYN
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
; FILE REFERENCE: 675002-2001
; CURRENT APPLICATION NUMBER: US/10/804,408
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: PCT/AU02/01281
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: AU PR 7749
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 172
; LENGTH: 2217
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
US-10-804-408-172

```

Query Match	9.3%	Score 647.6	DB 18	Length 2217
Best Local Similarity	59.5%	Pred. No. 1.2e-81		
Matches 1222	Conservative 0	Mismatches 794	Indels 38	Gaps 6
Qy	16	TGGCAATTATTGATGATGATAGCAGTTCGAAATCTTCGAAATCTTAAACAGTCATATACCAA	75	
Db	173	TAGCGATGATTCAAACAGTTGTGGTTATATTTCTCGAAGTTTGACATTTAACATTAAATTA	232	
Qy	76	ATGCTGATTTAAATTCG---TTCTCGAAATTTTATCATAAGTATGATGGTTCATTTATTTGGCAT	132	
Db	233	CTCCCAATTTTAAAGCAATAAAGATTTATTTGTTTGTCTATTTCATACATTTATATATGCTCT	292	
Qy	133	TTTTTATATCTCGTATGCCAGTTCGAATTTGAGTATAGAGGTAATCTCATAGAGTTTCAAA	192	
Db	293	TTTATCTTCTTGATTTTTTACAGAGACATTTTGGAGTCGTGGCTATCTTGAAGAGAGTTTAAAA	352	
Qy	193	AAACATTTAACTATAGTATATAATTTTGCAAATTTTCTTACGCCAGTATCAATTTTTGTGTGG	252	

Db 1424 TAAAGTTAGATGTTCAATATATCAATGATGCTATTTTGGTCAGATATTAAGATTATTC 1483  
Qy 1324 TAAAGACAGTGAAGTTGATTTGTTGAGAGGGAAGTAAGTAAAGTAT----- 1373  
Db 1484 TCCTAACACACTAAAGGTAGTCTTACTTGGGACAGGTCTAAGTAAAGGTAAAGTTGAAAG 1543  
Qy 1374 -----ATGAAAGTTTGGTTCGGTCTTCAAGGGGACATTTGACTCAGTGTATTT 1426  
Db 1544 GAATATATGAAATTTGTTCTGGTGGTTCAGTGGTGGTCACTAGCACACITGAACCT 1603  
Qy 1427 GTTAAACCGTTTGGGAAGGAAGAACGTTTGGGTGAACATTTGATGAAGAGATGC 1486  
Db 1604 TTTGAAACCCATTTGGGAAAGGAAGATAGTTTGGGTAAACCTTTGATGAAGAGATGC 1663  
Qy 1487 AAGAAGTCTTTGGAAGAAATGAAAAATGTATCCAGTTTCTTCCAAACAAATCGCAATCT 1546  
Db 1664 TAGGAGTATTTCTAAGAGAAAGATGTTATATCAATGCTTCTTCCAAACAAACCGTATGT 1723  
Qy 1547 CATTAATTTAGTGAATAATCTTTCTTAGCTTTTCABAAATTTAGCTGATGAGAACCGA 1606  
Db 1724 CAAAACTTGGTAAAAAATACTATCTAGCTTTTAAAGTCTTTAGAAAGAAAGACCGA 1783  
Qy 1607 TGTATTATTTCACTCGTGGCGCGTGTCTGCTCCCTTCTTTTACATCGGAAACATTT 1666  
Db 1784 TGTATCATATCACTGGTGGCGCTGTAGCAGTACCATTTCTTTATATTTGTAAGTTAT 1843  
Qy 1667 TGGAGCAAGACGATTTATATTTGAAGTATTTGATCGAGTTTAATAATCTACATTAACCTGG 1726  
Db 1844 TGGTTGTAAGACCGTTTATATAGAGGTTTTCGACAGGATAGATAAACCAACTTTGACAGG 1903  
Qy 1727 AAAACTAGTTTATCCCTGACAGATATTTTATTTGTTTCAGTGGGAAGAAATGAGAGGT 1786  
Db 1904 AAAATTAGTGATCTCTGTAACAGATAAATTTATTTGTTGAGTGGGAAGAAATGAAAAAGT 1963  
Qy 1787 ATATCCTAAATCTATTAACTTGGGAGTATTTTAAATGATTTTGTAAACAGTAGGAAT 1846  
Db 1964 TTATCCTAAGCAATTAATTAGAGGAAATTTTATGATTTTGTACAGTGGGACCA 2023  
Qy 1847 CATGAACACAGTTTAACTCGATGATGAAGAGATTTGATTTTATTTGAAAAAATGGAAGT 1906  
Db 2024 CATGAACAGCAGTTTCAACCGTCTTATTAAAGAGTTGATAGATTAAAGGACAGGTGCT 2083  
Qy 1907 ATAACCGACGAAATTTATTTCAACAGGATTTCTGACTATATTTCCAGATATTTGCAAG 1966  
Db 2084 ATTGATCAAGAAGTGTTCATTTCAACCGGTTACTCAGACTTTGAACCTCGAATTTGTCAG 2143  
Qy 1967 TATAAAAAATTTCTCAGTTTACAAAAGAAATGGAACAATATATTAACAATCAGAAGTAGTT 2026  
Db 2144 TGGTCAAAATTTCTCATATGATGATGATGAACTTTTACATGAAGAGCTGAGATTGTT 2203  
Qy 2027 ATTTGCCACGAGG 2040  
Db 2204 ATCACACATGGCG 2217

## RESULT 16

US-10-804-408-168  
; Sequence 168, Application US/10804408  
; Publication No. US20040253617A1  
; GENERAL INFORMATION:  
; APPLICANT: PANRONG, KONG  
; APPLICANT: GILBERT, GWENDOLYN  
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI  
; FILE REFERENCE: 675002-2001  
; CURRENT APPLICATION NUMBER: US/10/804,408  
; CURRENT FILING DATE: 2004-03-19  
; PRIOR APPLICATION NUMBER: PCT/AU02/01281  
; PRIOR FILING DATE: 2002-09-18  
; PRIOR APPLICATION NUMBER: AU PR 7749  
; PRIOR FILING DATE: 2001-09-19  
; NUMBER OF SEQ ID NOS: 182  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 168

; LENGTH: 2226  
; TYPE: DNA  
; ORGANISM: Streptococcus agalactiae  
; US-10-804-408-168  
Query Match 9.3%; Score 647.6; DB 18; Length 2226;  
Best Local Similarity 59.5%; Pred. No. 1.2e-81;  
Matches 1222; Conservative 0; Mismatches 794; Indels 38; Gaps 6;  
Qy 16 TGGCATTTATTTGATATGATAGCAGTTTGCATTTTCTCAATCTTTAAACAAGTCATATACCAA 75  
Db 182 TAGCGATGATTCACACAGTTTGGTTTATTTTCTGCAAGTTTGACATTAAACATTAATTA 241  
Qy 76 ATGCTGATTTAATATCG---TTCTGGAAATTTTATCATATATGATGGTTCATTTTTCAT 132  
Db 242 CTCCTCAATTTAAAGCAATAAAGATTTATTTGTTCTCTATTGATACATTTATTTGTCT 301  
Qy 133 TTTTATATCTCGTATGCCAGTTTGAATTTGAGTATAGAGTAACTCTGATAGAGTTTGAAA 192  
Db 302 TTTATCTTCTGATTTTACAGAGACTTTTGGAGTCTGGCTATCTTGAAGAGTTTAAAA 361  
Qy 193 AAACATTTAATCTAGTATATAATTTTGCATTTTCTTACGGCAGTATCATTTTGTGG 252  
Db 362 TGGTATTTGAAATACAGCTTTTACTATATTTTTCATATCAAGTTTCATTTTATTTTATTTA 421  
Qy 253 AGAATAATTTTCGCACCTTTCAGAGCTGGTGGCGGTGATTTTTCATTAATAAATCACTGTT 312  
Db 422 AAAACTCAITTTACACGACACGACTTTTCTCTTTTCTTTTATTTGCTATGAAATTCGATTT 481  
Qy 313 TGGTATACCTTATTTAAGTAAATTTTAAAGCAGTTTAAAGGATAGCTTTCTATTTCGACAA 372  
Db 482 TATTGATCTATTGAAATTCATTTTAAATAATTTATCGAAATAATTTCTTAGCTAAGTTT 541  
Qy 373 TCTATCAAAAAAGACGATTTCTAATTTACACCGCTGACGAGTGGGAAAAATATCAAGTTT 432  
Db 542 CACGAGATACCAAGTTTGTGATTAACGAATAAGGATTTCTTTATCAAAAAATGACCTTTA 601  
Qy 433 TATTGAAATCACATAAACAATTTCAAAAAATCTTGTGCAATTTGGTAGTTTATAGGTACAG 492  
Db 602 GGAATAAAT---ACGACCATTAATATATCGCTCTGTCATCTTTGGACTCTCTGAAAGG 658  
Qy 493 AATAGATATAAATTAATTTATCATTTACCTGCTCTATTATTCTGTTGGAAGAGCTATAGAGT 552  
Db 659 ATTGTTATGATTTGAAACATACTCGTTAAGGATTAATAACAAGATGCTCTTACTTTCAG 718  
Qy 553 TTTTCAACAAGGGAAGTGGTCGACCACTCTTTATAAATCTACCAAGTGAAGTTTATAG--- 609  
Db 719 AGTTAACCTGCTTAACTGTTGATCAAGCTTTTATTAACATACCACTTTGAATTTATTTGGTA 778  
Qy 610 -----ACGTAAAGCAATTCGTTTCAGATTTTTCAGATTTTGGTATTTAGTATGATGAGCGTTG 663  
Db 779 AATACCAAAATACAGATATTTAATTAATGACATTTGAAGCAATGGGAGTGAATTTGCAATGTTA 838  
Qy 664 ATATTAAATTCATTCGGTGTCTTACTGCGTTGAAAAACAAAAAATCCCAACTCTAGGTGACC 723  
Db 839 ATGTAGAGCACTTAGCTTTGATTAATATAGGAGAAAGCGAATCCAAACTTTTGAAGGAT 898  
Qy 724 ATAGCATTTGTAATCTTTTCCACAAAATTTTATAAGCCTTAGTCATATCATGATGAACGAC 783  
Db 899 ATAGTGTATACATATTTCTAGAAATTTCTATAAATATAGTCACTTTATAGCAAAACGAT 958  
Qy 784 TTTTGGATATACCTCGGAGCGGTAGTCGGGTAAATTTATTTGCTGATGATGATTTCTATTGTT 843  
Db 959 TTTTGGATATACCGGTGCTTATTTATAGTTTGTCTATGTCATATGTGCAATTTGTGGCAATTTTC 1018  
Qy 844 TAGTTCCTCAATTTTCTGAGAGATGGTGGACCGGCTATTTTGTCTCAGAAAAGTTCGAC 903  
Db 1019 TAGTTCACAAATCAGAAAAGATGGTGGACCGGCTATCTTTCTCAAAATATAGTAGTCTC 1078  
Qy 904 AGAATGACGCAATTTTACATTTCTACAAGTTTTCGATCGATGATGATTTGATGCTGAGGAGC 963  
Db 1079 GTAATGTTAGATTTTATAGATTCTATAAATTTAGATCAATGCGAGTAGATGCAACAA 1138







```
Db 779 AATACCAATACAGATATTATTAATGACATTGAAGCAATGGGAGTGATTTGCTCAATGTTA 838
Qy 664 ATATTAAATTCATTCGGTTTACTCGGTGTGAAAAACAAAAATCCAACTCCTGCTAGTGACC 723
Db 839 ATGTAGAGGACACTTAGCTTTGATAATATAGAGAAAAAGCGAATCCAAACTTTTGAAGGAT 898
Qy 724 ATAGCATTTGAACCTTTTCCACAAATTTTATAGCCCTAGTCTATATCATGATGAACGAC 783
Db 899 ATAGTGTTATACATAATCTATGAAATTTCTATAAATATATAGTCACCTTATAGCAAAACGAT 958
Qy 784 TTTTGGATATACTCGGAGCGTAGTCGGGTTAAATTTATTTGGGTATAGTTTCTATTTTGT 843
Db 959 TTTTGGATATACGGGTGCTATTATAGTTTGTCTATATGTGGCATTTGTGCGCAATTTTC 1018
Qy 844 TAGTTCCAAATTTTCGTAGAGATGGTGGACGGGCTATTTTGTCTCAGAAAACGAGTTGGAC 903
Db 1019 TAGTTCCGCAATCAGAAAAAGATGGTGGACGGGCTATCTTTTCTCAAAATAGAGTAGGTC 1078
Qy 904 AGAATGGACGCATATTTACATTTCTACAAGTTTCGATCGATGTATGTTGATGCTGAGGAGC 963
Db 1079 GTAATGGTAGGATTTTATAGTTCTTATAAATTCAGATCAATCGGAGTAGTGCAGAACAAA 1138
Qy 964 GCAAAAAGACTTCTCTCAGCCAAAACAGATGCAAGGGTGGGTATGTTTAAAAATGGAA 1023
Db 1139 TTAAGAAAGATTTATTAGTTTCACAATCAATGACAGGG-----CTAATGTTTAAAGTTAGA 1193
Qy 1024 AAACGATCCTAGAAATTTACTCCAAATGGACATTTTCATACGCCAAACAAACAAAGTTTGAAGAGT 1083
Db 1194 AGATGATCCTAGAAATTTACTAAAAATAGGAAAAATTTATTTCG-AAAAACAAAGCATAGATGAGT 1252
Qy 1084 TACCACAGTTTATAAATGTTTAAATTCGCGATATAGTCTAGTTGGTACACGCTCCACTA 1143
Db 1253 TGCCTCAATTTCTAATGTTTTTAAAGGTGATATGAGTTTGTAGTAGGAACACGCGCTCCCA 1312
Qy 1144 CAGTTGATGAATTTGAAAAATATACTCTGCTCAAAAGACAGATTCAGTTTAAAAACGAG 1203
Db 1313 CAGTTGATGAATATGAAAGTAAATTTCAACGCAAGAGCGAGCTTAGTTTAAAGCCAG 1372
Qy 1204 GGATTACAGGTCTCTGCGAGGTAGTGGTCTAGTATATATCAGACACTTCGACGACGTAG 1263
Db 1373 GAATCACTGGTTTGTGCAAAATATCTGGTAGAAAATATATTAATCTGATTTTGAATATCG 1432
Qy 1264 TTCGGTTGGACTTAGCATACATGATTAATTTGGACTATCTGGTCAGATATTAATAATTTAT 1323
Db 1433 TAAAGTTAGATGTTTCAATATATCAATGAATGGTCTATTGCTCAGATATTAAGATATTTC 1492
Qy 1324 TAAAGACAGTGAAGTTGTTATTTGTGAGAGAGGGAAGTAAAGTAAAGTAT----- 1373
Db 1493 TCCTAAACACTAAAGGTAGTTTACTTCGGGACAGAGCTTAGTAAAGGTAGGTTTGAAG 1552
Qy 1374 -----ATGAAAGTTTGTGGTGGTCTTTCAGGGGACATTTGACTCACTTGTATTT 1426
Db 1553 GAATATAATGAAAAATTTGCTGGTTGGTTCAAGTGGTGGTCACTAGCACACTTGAACCT 1612
Qy 1427 GTTAAACCGTTTGGAGGAGAGAACGTTTGGTGAACATTTGATTAAGAGGATGC 1486
Db 1613 TTTGAAACCCCAATTTGGGAAAAAGAGATAGGTTTTTGGGTAACTTTGATAAAGAGATGC 1672
Qy 1487 AAGAAGTCTTTTGAAGAAATGAAAAATGTATCCATGTTTACTTTCCAAACAAATCCCAATCT 1546
Db 1673 TAGGAGTATTTAGAGAGAGATTTGATATCATTTGCTTTCTTCCAAACAAACCGTAATGT 1732
Qy 1547 CATTAATTTAGTGAATAATCTTTCTTAGCTTTTCAAAATTTTACGTGATGAGAAACGAGA 1606
Db 1733 CAAAAACTTGGTAAAAAATACTATCTTAGCTTTTAAAGTCTCTTGAAGAAAGAACCCAGA 1792
Qy 1607 TGTATTATTATTCATCTGGTGGCGGTTGCTGTCGCCCTCTTTTACATCGAATAACTATT 1666
Db 1793 TGTATCATATCATCTGGTGGCGCTGTAGCAGTAGTACCATTCTTTTATATTTGGTAGTTATT 1852
Qy 1667 TGGAGCAAGACGATTTATATTTGAAGTATTTGATCGAGTTTAAATAAATCTACATTAACCTGG 1726
Db 1853 TGGTTGTGAAGACCGTTTATATATAGAGGTTTTCGACAGGATAGATAAACCAACTTTGACAGG 1912
```

## RESULT 18

```
US-10-804-408-164
; Sequence 164, Application US/10804408
; Publication No. US20040253617A1
; GENERAL INFORMATION:
; APPLICANT: FANRONG, KONG
; APPLICANT: GILBERT, GWENDOLYN
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
; FILE REFERENCE: 675002-2001
; CURRENT APPLICATION NUMBER: US/10/804,408
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: PCT/AU02/01281
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: AU PR 7749
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 164
; LENGTH: 2217
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
US-10-804-408-164
```

```
Query Match          9.2%; Score 646; DB 18; Length 2217;
Best Local Similarity 59.4%; Pred. No. 2e-81;
Matches 1221; Conservative 0; Mismatches 795; Indels 38; Gaps 6;
```

```
Qy 16 TGGCATTATTTGATATGATAGCAGTTGCAATTTCTGCAATCTTAAACAAGTCATATACCA 75
Db 173 TAGCGATGATTTCAACCGTTTGGTTTATTTTCTGCAAGTTTGACATTAACATTAATTA 232
Qy 76 ATGCTGATTTAAATCG---TTCTGGAAATTTTATCATATGATGGTTTCATTATTTTGCAT 132
Db 233 CTCCCAACTTTAAAGCAATAAAGATTTATTTGTTGTTCTTATTTGATACATTTATTTGCT 292
Qy 133 TTTTATATCTCGTATGCCAGTTGAAATTTGAGTATAGAGTAATCTGATAGAGTTTGAAA 192
Db 293 TTTATCTTTCTGATTTTACAGAGACTTTTGGAGTCGTGGCTATCTTGAAGAGTTTAAAA 352
Qy 193 AAACATTTAACTAGTATAATAATTTGCAATTTTCTTACGGCAGTATCATTTTGTGG 252
Db 353 TGGTATGAAATACAGCTTTTACTATATTTTCAATCAAGTTTCATTATTTTATTTTAA 412
Qy 253 AGAATAATTTCCGCACTTTCAAGACGTTGGTGGCGGTGATTTTTCACATTAATAAATTCGTT 312
Db 413 AAAAAGCTCAATTACACGACGACGCTTTTCTTTTACTTTTCTTTTATGCTATGAATTCGATTT 472
```



Matches 1220; Conservative 0; Mismatches 796; Indels 38; Gaps 6;	
Qy	16 TGGCAATTTATGATAGAGTTCGAATTTCTGCAATCTTAAACAAGTCATATACCAA 75
Db	
Db	173 TAGCGATGATTCAAACAGTTGTGTTATGTTCTGTAAGTTTGCACATTAACATTAATCA 232
Qy	76 ATGCTGATTTAAATCG--TTCTGGAATTTTATCAATAGATGGTTTCATTTTTCAT 132
Db	
Db	233 CTCCCAATTTAAAGCAATAAGATTTATTTGTTGTTCTATTGATACATATATTGTCT 292
Qy	133 TTTTATATCTCGTATGCCAGTTGAAATTGAGTATAGAGTAACTCTGATAGAGTTTGAAA 192
Db	
Db	293 TTTATCTTTCTGATTTTACAGAGCTTTTGGAGTCTGCTATCTTGAAGAGTTTAAAA 352
Qy	193 AAACATTTAACTATAGTATTAATATTTTGCATTTTCTTACGGCAGTATCAATTTTGTGG 252
Db	
Db	353 TGGTATTTGAAATACAGCTTTTACTATATTTTCAATCAAGTTTCATTTATTTTATTTT 412
Qy	253 AGATTAATTTTCGCACTTTCAAGAGCTGGTCCGCTGATTTTTCACATTAATAACTTCGTTT 312
Db	
Db	413 AAAAATCTTTTACAAACAGCAGCTTTTCTTTTACTTTTATTTGCTATGAAATTCGATTT 472
Qy	313 TGGTATACCTATTTAACGTAATTAATAAGCAGTTTAAAGGATAGCTTTCTATTTTCGACAA 372
Db	
Db	473 TATTATATCTATTGAAATTCATTTTAA--AAATATTCGAAATATTTCTTACGCTAAGT 529
Qy	373 TCTATCAAAAAGAGCAGTATCTAATTAACAAGCTGTAACAGATGGGAAATATGCAAGTTT 432
Db	
Db	530 TTTCAAGAGATACCAAAGTTGTTTGTATACGAAATAGGATTTCTTTTCAAAAATGACCT 589
Qy	433 TATTGTAATCACATAACAAATTCAAAATAATCTTGTGATCGGTAGTTTATAGTACAG 492
Db	
Db	590 TTAGGAACAAATACGACCATTAATATATCGCTGCTGTATCTTGGACTCTCTGAAAGG 649
Qy	493 AAATAGATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 552
Db	
Db	650 ATTGTTATGATTGGAACATTAATCTGTTAGGATAATAACAAGATGCTTTACTTTCAG 709
Qy	553 TTTCAACAAGGAAGTGGTGGACACGCTTTTATTAATTAATTAATTAATTAATTAATTA 609
Db	
Db	710 AGTTAACTGCTTAATCTGATCAAGCTTTTATTAACATACCCATTTGAATTTATTTGGTA 769
Qy	610 -----AGTTAAAGCAATTCGTTTTCAGATTTTGGATTTGTTAGGTTATGATGAAGCGTTG 663
Db	
Db	770 AATACCAATACAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 829
Qy	664 ATATTAATTCATTCGGTTTACTCGGTTTGAACAAACAAACAAATCCAACTGCTAGTGCACC 723
Db	
Db	830 ATGTAGAGGCACTTAGCTTTGATTAATATAGGAGAAAGCGAATCCAACTTTTGAAGAT 889
Qy	724 ATAGCATTTGAATCTTTTCCAAATTTTATAAGCCCTAGTCAATCATGATGAACGAC 783
Db	
Db	890 ATAGTGTATTACATATCTATGAAATTTCTATATAATATAGTCACTTTATAGCAAAACGAT 949
Qy	784 TTTTGGATATCTCGGAGCGTAGTTCGGTTTAAATTTTGTGTTAGTATGATTTCTATTTTGT 843
Db	
Db	950 TTTTGGATATCATGGGTGCTATTATAGGTTTGTCTCATATGCGCAATTTTGTGCGCAATTTTC 1009
Qy	844 TAGTTTCAAATTTTCTGATAGATGCTGGACCGGCTATTTTGTCTCAGAAACGAGTTCGAC 903
Db	
Db	1010 TAGTTTCCGAATCAGAAAGATGGCGGACCGGCTATCTTTTCTCAAAATATAGATAGTCT 1069
Qy	904 AGAATGGACGCATATTTTACATTTTACAGTTTTCGATCGATGATGTTGTTGATGCTGAGGAC 963
Db	
Db	1070 GTAATGGTAGGATTTTATGATTTCTATAAATTCAGATCAATGCGAGTAGTAGCAACAAA 1129
Qy	964 GCAAAAAGACTTCTCAGCCAAACCAAGATGCAAGGTTGGGTATGTTTAAATGGGAA 1023
Db	
Db	1130 TTAAGAAAGATTTTATAGTTTACCAATCAAAATGACAGGG-----CTAAATGTTTAAAGTTTGA 1184
Qy	1024 AAACGATCCTAGAATTTACTCCAAATTTGGACATTTTCATACGCAAAAACAAAGTTTACAGGAT 1083
Db	
Db	1185 AGATGATCCTAGAAATTTACTAAATAGGAAATTTTATTCG-AAAAACAAGCATAGATGAAT 1243

Qy	1084 TACCACAGTTTATATATGTTTAAATTGGCGATATAGTCTAGTTGGTACACGTTCCACCTA 1143
Db	
Db	1244 TGCCTCAATTTCTATATATGTTTAAAAAGTGTATATGATTTAGTAGGAACACGCGCTCCCA 1303
Qy	1144 CAGTTGATCAATTTTGAATAATATACCTCGTCTCAAAAGAGACGATTTGAGTTTAAACCAG 1203
Db	
Db	1304 CAGTTGATGAATATGAATAATATTAATCAACGCAAGAGCGCCCTTAGTTTAAAGCCAG 1363
Qy	1204 GGAATACAGCTCTCTCGCAGGTTAGTGGTCTAGTAAATATCAACAGACTTTTCGACAGCTAG 1263
Db	
Db	1364 GAATCACTGGTTTGTGCGCAATATCTGTTAGAAATATATTAATTAATCTGATTTGATGAATCG 1423
Qy	1264 TTCGGTTGACCTTAGCATATCAATTTGATTAATTTGGAATCTCTGGTCAGATATTTAAATTTTAT 1323
Db	
Db	1424 TAAAGTTAGATGTTTCAATATATCAATGAATGGTCTATTTGGTCAGATATTTAAGATTTATC 1483
Qy	1324 TAAAGACAGTGAAGTTGTTATTTGTTTCAGAGAGGGAAGTAAAGTAAAGTAT----- 1373
Db	
Db	1484 TCATAACACTTAAAGGTTAGTTTACTCGGACAGGAGCTAAGTAAAGGTAAGGTTTGAAG 1543
Qy	1374 -----ATCAAAAGTTTGTTCGTTTGGTTCGTTTCTTTCAGGGGACATTTTGACTCAGTTGTTATTT 1426
Db	
Db	1544 GAATATAATGAATAATTTTGTCTGGTTCAAGTGGTGGTCACTAGCACACTTTGAACCT 1603
Qy	1427 GTTAAAAACCGTTTGGAGAGAGAAACGTTTTTGGGTAAACATTTTGGTAAAGAGATGC 1486
Db	
Db	1604 TTTGAAACCCATTTTGGGAAAAAGAGATAGTTTTTGGGTAAACCTTTTGAATAAGAGATGC 1563
Qy	1487 AGAAGCTCTTTGAAGAATGMAAAATGATCCATGTTACTTTTCCACCAATCGCAATCT 1546
Db	
Db	1664 TAGGATATTTCTAAGAGAGAGATTTGTATATCATCTCTTTTCCCAACCAACCGTAAATGT 1723
Qy	1547 CATTAATTTAGTGAAAAATACCTTTCTTAGCTTTTCAAAATTTTACGTGATGAGAAACCGA 1606
Db	
Db	1724 CAAAACTTGGTAAAAAATACTATTCTAGCTTTTAAAGTCTTTAGAAAAGAGAACCGA 1783
Qy	1607 TGTATATTTTCACTGGTGGCGCGTGTGTCCTCTCTTTTACATCGGAAAACTATT 1666
Db	
Db	1784 TGTATCATATCATCTGGTCCGCTGTAGCAGTACCATTCTTTTATATTTGTTAAGTTTAT 1843
Qy	1667 TGGAGCAAGACGATTTATATTTGAAGTATTTGATCGAGTTTAAATAATCTCATTTAACTGG 1726
Db	
Db	1844 TGGTTGAAGACCGTTTATATAGAGGTTTTCGACAGATAGATAAACCAACTTTGACAGG 1903
Qy	1727 AAAATAGTTTATCCCGTAAACAGATATTTTATTTGTTTCAGTGGGAAGAAATGAAGAAGT 1786
Db	
Db	1904 AAAATAGTGTATCTGTAAACAGATTAATTTATTTGTTTCAGTGGGAGAAATGAAGAAAT 1963
Qy	1787 ATATCCTAAATCTATTAACTTTGGGAGTATTTTTTAAATGATTTTGTAAACAGTAGGAAT 1846
Db	
Db	1964 TTATCCTAAGGCAATTAATTTTAGAGGAAATTTTAAATGATTTTGTACAGTGGGACCA 2023
Qy	1847 CATGAACACAGTTTAAATCGATTTGAAGAAGATTTGATTTTAAAGAAATTTGAAGAAATTTGAAGT 1906
Db	
Db	2024 CATGAACAGAGTTTCAACCGCTTTATTAAGAAAGTTGATAGATTAAGAAAGGACAGGTGCT 2083
Qy	1907 ATAAACCGACGAAATATTTTAAATTTCAACAGGATATTTCTGACTATATTTCCAGATATTTGCAAG 1966
Db	
Db	2084 ATTGATCAGAGAGTGTTCATTCAACCGGTTACTCAGACTTTGAACTTCAGAAATTTGTCAG 2143
Qy	1967 TATAAAAAATTTCTAGTTTCAAGAAATGGAACAATATATTTAAACAAATCAGAGTAGTT 2026
Db	
Db	2144 TGGTCAAAATTTCTCTCATATGATGATGAACTCTTACATGAAAGAGCTGAGATTGTT 2203
Qy	2027 ATTTGCCACGAGG 2040
Db	
Db	2204 ATCACACATGGCGG 2217

Publication No. US20040253617A1  
; GENERAL INFORMATION:  
; APPLICANT: FANRONG, KONG  
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI  
; FILE REFERENCE: 675002-2001  
; CURRENT APPLICATION NUMBER: US/10/804,408  
; PRIORITY FILING DATE: 2004-03-19  
; PRIOR APPLICATION NUMBER: PCT/AU02/01281  
; PRIOR FILING DATE: 2002-09-18  
; PRIOR APPLICATION NUMBER: AU PR 7749  
; PRIOR FILING DATE: 2001-09-19  
; NUMBER OF SEQ ID NOS: 182  
; SOFTWARE: Patent in version 3.2  
; SEQ ID NO 170  
; LENGTH: 2226  
; TYPE: DNA  
; ORGANISM: Streptococcus agalactiae  
US-10-804-408-170

Query Match 9.2%; Score 644.4; DB 18; Length 2226;  
Best Local Similarity 59.4%; Pred. No. 3.4e-81;  
Matches 1220; Conservative 0; Mismatches 796; Indels 38; Gaps 6;  
Qy 16 TGGCAATTAATGATAGTAGAGTTGCAATTTCTGCAATCTTAACAGTCATATACCAA 75  
Db 182 TAGCGATGATTCAAACCGTTGTGGTTTATTTTCTGCAAGTTTGACATTAACATTAATTA 241  
Qy 76 ATGCTGATTAATTCG---TTCTGGAATTTTATCATATGATGGTTCAATTTTGGCAT 132  
Db 242 CTCCCAACTTTAAAGCAATAAGATTTATTTGTTGTTCTATGATACATTAATATTTGCT 301  
Qy 133 TTTTATATCTCGTATGCGAGTTGAAATTTGAGTATAGAGTAATCTGATAGAGTTTGAAA 192  
Db 302 TTATCTTCTGATTTTACAGAGACTTTTGGAGTCGTGGCTAICTTGAAGAGTTTAAA 361  
Qy 193 AATCAATTAATAGTATATATATTTGCAATTTTCTTACGCGATGATCATTTTGTGG 252  
Db 362 TGGTATTAATGAAATACAGCTTTTACTATATTTTATCATATCAAGTTTCATTTATTTT 421  
Qy 253 AGAATTAATTCGCACTTTCAAGAGTGTGGCGGTATTTTACATTAATAAATTCGTTT 312  
Db 422 AAAAATCTTTTACAAACGACGACTTTCCTTTTACTTTTATTAATGATTAATTCGATTT 481  
Qy 313 TGGTATACCTATTTAAACGTAATTAATAGCAGTTTAAGGATAGCTTCTATTTTCGACAA 372  
Db 482 TATTATATCTATGAAATTCATTTTAAATATTAATTCGAAATATCTTACCGTAAGTTT 541  
Qy 373 TCTATCAAAAAAGACGATTTCTAATTAACCGCTGAACGATGGGAAATATGCAAGTTT 432  
Db 542 CACGAGATACCAAAGTTGTTTGTATACGAATAAGGATTTCTTTATCAAAATAGACCTTTA 601  
Qy 433 TATTGATACATAAACAATTTCAAAAAATCTTGTGCAATGCGTAGTTTGTAGTACAG 492  
Db 602 GGAATAAAT---ACGACCAATTAATATATCGCTGTCTGTATCTGATCTCTCTGAAAGG 658  
Qy 493 AAATAGATAAATAATTTATCATTTACCGCTCTATTTCTGTGGAAGAACTATAGAT 552  
Db 659 ATTGTTATGATTTGAAACATAACTCGTTTAAAGATAATAAACAAGATGCTCTTACTTCAG 718  
Qy 553 TTTCAACAGGGAAGTGTGCGACCACTCTTTTATAAATCTACCAAGTGAGTTTGTAG--- 609  
Db 719 AGTTAACCTGCTTAACCTGTGATCAAGCTTTTATTAACATACCCATTTGAATTTTGTGA 778  
Qy 610 -----ACGTAAAGCAATTCGTTTCAGATTTTGTAGTTGTTAGGTTATGATGTAAGCGTTG 663  
Db 779 AATACCAATTAACAAGATATTAATGACATTAAGCAATGGGAGTGATTTGTCATGTTA 838  
Qy 664 ATATTAATTCATTCGTTTCTGCTGGAACCAAAAAATCCAACTGCTAGGTGACC 723  
Db 839 ATGTAGAGGCACTTAGCTTTGATTAATATAGAGAAAGCGAATCCAACTTTTGAAGGAT 898  
Qy 724 ATAGCAATGTAACCTTTTCCACAAATTTTATAAGCCCTAGTCATATCATGATGAACGAC 783

Db 899 ATAGTGTATTACATATTTCTATGAATTTCTATAAATATAGTCACCTATAGCAAAACGAT 958  
Qy 784 TTTTGGATATATCTCGAGCGGTAGTCGGGTAAATTTATTTGGGTATAGTTCTTATTTGT 843  
Db 959 TTTTGGATATACCGGTGCTATTAATAGTTTGGCTCATATGTCGCAATTTGGCAATTTTC 1018  
Qy 844 TAGTTCCAATTTATTCGTAGAGATCGTGACCGGCTATTTTGGCTCAGAAAACGAGTTGAC 903  
Db 1019 TAGTTCACAAATCAGAAAAGATGTCGACCGGCTATCTTTTCTCAAAATAGAGTAGGTC 1078  
Qy 904 AGAATGACGCAATTTTACATTTCTCAAGTTTCGATCGATGATGTTGCTGCTAGGAGC 963  
Db 1079 GTAATGGTAGGATTTTAGATTTCTATAAATTCAGATCAATCGAGTAGATGCAAGACAAA 1138  
Qy 964 GCAAAAAAGACTGCTCAGCCCAAAACAGATGCAAGGGTGGGTATGTTTAAAAATGGGAA 1023  
Db 1139 TTAAGAAAGATTTATTTAGTTTCAATCAATCAATGACAGGG-----CTAATGTTTNAAGTTAGA 1193  
Qy 1024 AAACGATCCTAGAAATTAATCTCAATTTGGACATTTTATAGCGAAAACAAAGTTTACAGAGT 1083  
Db 1194 AGATGATCCTAGAAATTTACTAAAAATAGGAAAATTTATTTCG-AAAAACAAGCATAGATAGT 1252  
Qy 1084 TACCACAGTTTATTAATGTTTAAATTTGCGGATATAGTCTAGTTGGTACACGTCACCTA 1143  
Db 1253 TGCCTCAATTTCTAATGTTTAAAGGTGATATGAGTTTATAGTAGGACACGCGCTCCCA 1312  
Qy 1144 CAGTTGATGAATTTGAAAATATATCTCTGCTCAAAAGAGACGATTTAGTTTAAACCCAG 1203  
Db 1313 CAGTTGATGAATTTGAAAAGTATTAATTTCAACGAGAGAGCGCTTGTAGTTTAAAGCCAG 1372  
Qy 1204 GGATTACAGGTCCTCGCAGGTTAGTGGTCTAGTATATATACAGACTTTCGACGACGTAG 1263  
Db 1373 GAATCACTGGTTTGTGCAAAATATCTGTGAGAAATATATTTACTGATTTTGTGATGAATCG 1432  
Qy 1264 TCCGTTGGACTTAGCATACATGATTAATTTGGACTATCTGTCAGATTAATAAATTTTAT 1323  
Db 1433 TAAAGTTAGATGTTTCAATATATCAATGAATGCTCTATTTGGTCAGATTAAGATTTATTC 1492  
Qy 1324 TAAAGACAGTGAAAGTTGTTGTTGAGAGAGGGAAGTAAGTAAAGATAT----- 1373  
Db 1493 TCCTAAACACTAAAGTAGTCTTACTTGGGACAGGTGCTTAAGTAAAGTTAGTTTGAAG 1552  
Qy 1374 -----ATGAAAGTTTGTGGTTCGTTCTTACGGGGGCAATTTGACTCACTGTATTT 1426  
Db 1553 GAATATTAATGAAATTTGTTGGTTGGTTCAAGTTGGTGTCTATCTAGCACACTTGAACCT 1612  
Qy 1427 GTTAAACCGTTTGGAGGAAAGAACGTTTGGTGTAAACATTTGATTAAGAGGATGC 1486  
Db 1613 TTTGAAACCCATTTTGGAAAAGAGATAGGTTTGGGTAACTTTGATAAAGAGATGC 1672  
Qy 1487 AAGAGTCTTTTGAAGATGAAAATGATCCATGTTTACTTTCCACAAATCGCAATCT 1546  
Db 1673 TAGGATATTTCTAAGAGAGATTTGATATCATTTGCTTCTTTCCAAACCCGTAATGT 1732  
Qy 1547 CATTAATTTAGTGAATAATCTTTCTTAGCTTTTCAAAATTTTACGTGATGAGAACACAGA 1606  
Db 1733 CAAAACTTGGTAAAAATATCTATTTAGCTTTTAAAGTCTTTAGAAAAGAAAGNCCAGA 1792  
Qy 1607 TGTATTATTTTCACTCGTGGCGCGTGTGTCCTTCTTTTACATCGGAAACATATT 1666  
Db 1793 TGTATCATATCATCTGTCGCGCTGTAGCAGTACCATTATTTTATATTTTATATTTGTAAGTTATT 1852  
Qy 1667 TGGAGCAAGACGATTTTATTTGAAGTATTTGATCGAGTTAATAAATCAATTAACCTGG 1726  
Db 1853 TGGTTGTAAGACCGTTTATATAGAGGTTTTCGACAGGATAGATAAACCCTTTTGACAGG 1912  
Qy 1727 AAAACTAGTTTATCCCGTAAACAGATATTTTATTTGTTTCAAGTGGGAGAAATGAAGAAGT 1786  
Db 1913 AAAATAGTATCTGTTAAACAGATAAATTTATTTGTTTCAAGTGGGAGAAATGAANAAGT 1972  
Qy 1787 ATATCTTAAATCTATTAACTTTGGGAGGTATTTTAAATGATTTTGTAAACAGTAGGAACT 1846

Db 1973 TTATCCTAAGCAATTAATTTAGGAGGAAATTTTAAATGATTTTTTGTGCACAGTGGGACA 2032  
Qy 1847 CATGAACAACAGTTTAAATCAATGATTAAGAGAGATTGATTTATTTGAAAAAATGGAAGT 1906  
Db 2033 CATGAACAGAGTTTCAACCGCTTATTTAAAGAGTTGATAGATTTAAAGGGACAGATGCT 2092  
Qy 1907 ATAACCCAGCAAAATATTTATTTCAACAGAGGATTTCTGACTATATTTCCAGAAATTTGCAAG 1966  
Db 2093 ATTGATCAGAAGTTGTCATTTCAACCGGTTACTCAGACTTTGAACTCGAATTTGTCAG 2152  
Qy 1967 TATAAAAAATTTCTCAGTTCAAGAAATGGAACAATATATTTAAACAAATCAGAAGTAGTT 2026  
Db 2153 TGGTCAAAATTTCTCTCATATGATGATGATGAATCTTTACATGAAGGAAGCTGAGATTTGT 2212  
Qy 2027 ATTTGCCACGGAG 2040  
Db 2213 ATCACATGGCGG 2226

RESULT 21  
US-10-804-408-165  
; Sequence 165, Application US/10804408  
; Publication No. US20040253617A1  
; GENERAL INFORMATION:  
; APPLICANT: FANRONG, KONG  
; APPLICANT: GILBERT, GWENDOLYN  
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI  
; FILE REFERENCE: 675002-2001  
; CURRENT APPLICATION NUMBER: US/10/804,408  
; CURRENT FILING DATE: 2004-03-19  
; PRIOR APPLICATION NUMBER: PCT/AU02/01281  
; PRIOR FILING DATE: 2002-09-18  
; PRIOR APPLICATION NUMBER: AU PR 7749  
; PRIOR FILING DATE: 2001-09-19  
; NUMBER OF SEQ ID NOS: 182  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 165  
; LENGTH: 2225  
; TYPE: DNA  
; ORGANISM: Streptococcus agalactiae  
US-10-804-408-165

Query Match 9.18; Score 638.8; DB 18; Length 2225;  
Best Local Similarity 59.58; Pred. No. 2.1e-80;  
Matches 1223; Conservative 0; Mismatches 792; Indels 39; Gaps 7;

Qy 16 TGGCAATTAATTTGATGATAGCAGTTGCAATTTCTGCAATCTTAAACAAGTCATATACCAA 75  
Db 182 TAGCGATGATTCAAACAGTTGTGGTTTATTTTCTGCAAGTTTGACATTAACATTAATTA 241  
Qy 76 ATGCTGATTTAAATCG---TTCGGAAATTTTATCATATAATGATGGTTTCATTTTTCGAT 132  
Db 242 CTCCCAATTTTAAAGCAATAAAGATTTATTTGTTGTTCTATTGATACATTTATTTGTCT 301  
Qy 133 TTTTATATCTCGATCGCAGTTCAATTTGAGTATAGAGTAATCTCATAGAGTTTGAAA 192  
Db 302 TTTAATCTTCGATTTTACAGAGACTTTTGGAGCTGGTGGCTATCTTTGAAGAGTTTAAAA 361  
Qy 193 AAAACATTAACATAGTATATAATTTGCAATTTTCTTACGGCAGTATCAATTTTGTGG 252  
Db 362 TGGTATTTGAATAACAGCTTTTACTATATTTTCATATCAAGTTTCATTTATTTATTTTA 421  
Qy 253 AGAATAATTTGCACTTTCAAGAGCTGGCGGTGATTTTACATAATTAATTAACCTCGTTT 312  
Db 422 AAAACTCAATTTACAATGACAGACTTTTCCCTTTTTCCTTTTATTTGCTATGAAATTCGATTT 481  
Qy 313 TGGTATACCTATTTAACGTAATTAATTAAGCAGTTTAAGGATAGCTTTCTATTTCGACAA 372  
Db 482 TATTATATCTATTTGAATTCATTTTTTAAATAATTAATCGAAATAATTTCTTACGCTAAGTTT 541  
Qy 373 TCTATCAAAAAGACAGCTTCTAATTAACAACGGCTGAACGATGGGAAATATGCAAGTTT 432  
Db 542 CACGAGATACCAAGTTGTTTGTGATAACGAATAAGGATTTCTTTATCAAAAATGACCTTTA 601

Qy 433 TATTTGAATCACAATAACAATAATCTTGTTCGATTTGGTAGTTTATAGGTACAG 492  
Db 602 AGAATAAAT---ACGACCATAATATATCGCTGTCTGTATCTTTGGGACTCTCTGAAAGG 658  
Qy 493 AAATAGATAAAATTAATTTATCATTTACCCTCTATTATTCTGTGGGAAGACTATAGAGT 552  
Db 659 ATTGTTATGATTTGAAACATAACTCGTTAAGGATATAAACAAGATGCTCTTACTTTCAG 718  
Qy 553 TTTCAAAGAGGGAAGTGGCGACCAGCTCTTTATAAATCTTACCAAGTAGTTTTTAG--- 609  
Db 719 AGTTAACCTGCTTAACATGTTGATCAAGCTTTTATTAACTACCACTTGAATTTATTTGGTA 778  
Qy 610 -----ACSTAAAGCAATTCGTTTCAGATTTTCAGTTTGTAGTATTGATGTAAGCGTTG 663  
Db 779 AATACCAAAATACAAGATATTTAATTAATGACATTTGAAGCAATGGGAGTGAATGTCAATGTTA 838  
Qy 664 ATATTAAATTCATTCGGTTTTTACTCGCTTGAAAAACAAAAAATCCAACTGCTAGGTGACC 723  
Db 839 ATGTAGAGGCACTTAGCTTTTGATATATAGGAGAAAGCGAATCCAACTTTTGAAGGAT 898  
Qy 724 ATAGCAATTTGTAATTTTCCACAAATTTTATTAAGCCTAGTCAATATCATGATGAAAAACGAC 783  
Db 899 ATAGTGTATTACATATTTCTATGAAATTTCTATAAATATAGTCACTTATAGCAAAACGAT 958  
Qy 784 TTTTGGATATACTCGGAGCGGTAGTCCGGTTAAATTTATTTTGTGTTAGTATTCTTATTTGT 843  
Db 959 TTTTGGATATCATGGGTGCTATTTATAGTTTGTCTCATATGTGGCATTTTGGCAATTTTC 1018  
Qy 844 TAGTTCCAATTTATTCGTAGAGATGGTGGACCGCTATTTTGTCTCAGAAACGAGTTGGAC 903  
Db 1019 TAGTTCGCAATCAGAAAGATGGTGGACCGCTATCTTTCTCAAAATAGAGTAGGTC 1078  
Qy 904 AGAATGCGACGCAATTTTACATTTCTACAAGTTTCGATCGATGTATGTTGATGCTGAGGAGC 963  
Db 1079 GTAATGCTAGGATTTTGTAGATT-TATAAATTCAGATCAATGCGAGTAGATGCAAGACAA 1137  
Qy 964 GCAAAAAAGACTTGTCTCAGCCAAAACCAAGATGCAAGGGTGGGTATGTTTTAAATGGGAA 1023  
Db 1138 TTAAGAAAGATTTTATTTAGTTTCAAAATCAAAATGACAGG-----CTAATGTTTAAAGTTAGA 1192  
Qy 1024 AAACGATCCTAGAAATTTACTTCCAATTCGACATTTTATACGCAAAAACAAGTTTATAGACAGT 1083  
Db 1193 CGATGATCCTAGAAATTTCTAAATAGAGAAATTTATTCG-AAAAACAGCATAGATGAGT 1251  
Qy 1084 TACCACAGTTTATAATGTTTAAATTTGGCGATATGAGTCTAGTTGGTACACGCTCCACCTA 1143  
Db 1252 TGCCTCAATTTCTATAATGTTTAAAGGTCATATGAGTTTAGTAGGAACACGCGCTCCCA 1311  
Qy 1144 CAGTTGATGAATTTGAAAAATATATCTCTGGTCAAAAGAGACGATGAGTTTAAACCAG 1203  
Db 1312 CAGTTGATGAATATGAAGAAGTATAATTCAAACGAGAAAGCGACGCTTAGTTTAAAGCAG 1371  
Qy 1204 GGAATACAGTCTCTCGCAGGTTTAGTGGTCTAGTAAATATACAGACTTCGACAGCTAG 1263  
Db 1372 GAATCACTGTTTGTGCGCAATATCTGGTAGAAATTAATTAATCTGATTTTGATGAAATCG 1431  
Qy 1264 TTCGTTTGGACTTTAGCATACATTTGATAATTTGGAATATCTGTTGTCAGATATTTAAATTTAT 1323  
Db 1432 TAAAGTTAGATGTTCAATATATCAATGAATGGTCTATTTGGTCAGATATTTAAGATTTATTC 1491  
Qy 1324 TAAAGACAGTGAAGTTGTTATTTGTTAGAGAGGGAAGTAGTAAGTAAAGTAAAGTAT- 1373  
Db 1492 TCCTAACTTAAAGGTAAGTCTTACTTTGGGACAGGAGCTAAGTAAAGTAAAGTTTGAAG 1551  
Qy 1374 -----ATGAAAGTTTGTGTTGCTGCTCTTTCAGGGGACATTTGACTCAGTTTGTATTT 1426  
Db 1552 GAATATAATGAATTTTGTGTTGCTGTTCAAGTGGTGGTCACTTAGCACCTTTGAACCT 1611  
Qy 1427 GTTAAACCGTTTTTGGAGGAAGAAAGACGTTTTTTGGGTAAACATTTTGTAAAGAGGATGC 1486  
Db 1612 TTTGAAACCCATTTGGGAAAAAAGAGATAGGTTTTTGGGTAAACCTTTTGATAAAGAGATGC 1671

```
Qy 1487 AAGAGTCTTTTGAAGAAATGAAAAATGATATCCATGTTTACCTTCCAAACAAATCCCAATCT 1546
Db 1672 TAGGAGTATCTTAAGAGAGAGATGTTATATCAITGCTTCTTTCCAAACAAACCGTAATGT 1731
Qy 1547 CATTAATTTAGTGAATAATCTTTCTTAGCTTTTCAAAATTTTACGTGATGAGAACACAGA 1606
Db 1732 CAAAACTTGGTAAAAAATATCTTCTAGCTTTTAAAGTCTTTAGAAAGAACACACAGA 1791
Qy 1607 TGTTATTATTTTATCTGCTGGCGGCTTGTCTGCCCTTCTTTTACATCGGAAATCTATT 1666
Db 1792 TGTTATCATATCAITCTGCTGGCGCTGTAGCAGTACCATTTCTTTTATATTGTAAGTATT 1851
Qy 1667 TGGAGCAAGACGATTTATATTTAGATTTTGTATGTCAGTAAATCTATTAATCTAACTGG 1726
Db 1852 TGGTTGTAAGACCGTTTATATAGAGGTTTTCGACAGGATAGATAACCACTTTGCACAGG 1911
Qy 1727 AAAAAGTATTTATCCGTAACAGATATTTTATTTGTTTCCAGTGGGAAGAAATGAAGAAGT 1786
Db 1912 AAAATTAGTGTATCTGTAAACAGATAAATTTATTTGTTTCCAGTGGGAAGAAATGAAGAAGT 1971
Qy 1787 ATATCCTAAATCTATTAACTTTGGGAGTATTTTAAATGATTTTGTAAACAGTAGGAAT 1846
Db 1972 TTATCCTAAGGCAATTAATTTAGGAGGAATTTTAAATGATTTTGTACAGTGGGACA 2031
Qy 1847 CATGAACAAAGTTTAAATCGATGATGATAAAGAGATTTGATTTTAAAGAAATGGAAGT 1906
Db 2032 CATGAACAGCAGTTTCAACCGTCTTATTAAAGAGTTGATAGATTAAAGAGGACAGGTGCT 2091
Qy 1907 ATACCCGACGAATATTTTAAACAGAGATTTTCAAGATTTTGTGACTATATTTCCAGATATTGCAAG 1966
Db 2092 ATTGATCAAGAAAGTGTTCATTTCAACCGGTTTACTCAGACTTTGAACTTCAAGATTGTGAG 2151
Qy 1967 TATAAAAAATTTCTCAGTTTACAAAGAAATGAAACAAATATATTAACAAATCAGAAATGAGTT 2026
Db 2152 TGCTCAAAATTTCTCATATGATGATGATGAATCTTACATGAAGAGTGTGATTTGTT 2211
Qy 2027 ATTTGCCACGAGG 2040
Db 2212 ATCACATGCGCG 2225
```

## RESULT 22

```
US-10-182-960-1
; Sequence 1, Application US/10182960
; Publication No. US20030186392A1
; GENERAL INFORMATION:
; APPLICANT: Tremphy, Janine, et al.
; TITLE OF INVENTION: Biopolymer Thickner
; FILE REFERENCE: 64376
; CURRENT FILING DATE: 2002-08-03
; PRIOR FILING DATE: 2002-08-03
; PRIOR APPLICATION NUMBER: US/10/182,960
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: PCT/US01/03404
; PRIOR FILING DATE: 2000-10-16
; PRIOR APPLICATION NUMBER: 60/241,098
; PRIOR FILING DATE: 2000-02-02
; PRIOR APPLICATION NUMBER: 60/179,888
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 1
; LENGTH: 6850
; TYPE: DNA
; ORGANISM: Lactococcus lactis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (174)..(488)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (528)..(977)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1020)..(1796)
; FEATURE:

; NAME/KEY: CDS
; LOCATION: (1809)..(2501)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2618)..(3307)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3332)..(4015)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (4022)..(4468)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (4974)..(5678)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (5687)..(6778)
; US-10-182-960-1

Query Match 6.1%; Score 428.4; DB 16; Length 6850;
Best Local Similarity 58.9%; Pred. No. 1.3e-50;
Matches 849; Conservative 0; Mismatches 581; Indels 12; Gaps 6;

Qy 861 AGAGATGCTGACCGGCTATTTTGTCTCAGAAACGAGTTGACAGAAATGGACGATATTT 920
Db 3518 AAAGATCAAGGGCCAAATGTTCTATATAACAAACGCTATGTTAAATAATGGTAAATTTT 3577
Qy 921 ACATTCTACAAGTTTTCGATCGATGTATGTTGATCTCGAGGAGCGCAAAAAGACTTGTCT 980
Db 3578 TATATTTTGAATTTTAGAACAAATGATTTCTTAATCCGAGCA--GTAICTAGAACTTAATC 3635
Qy 981 AGCCAAAACACAGATGCAAGGGTGGTATGTTTAAAAATGGGAAAAACGATCCTAGAAATTA 1040
Db 3636 CAGATGTTAAAGCTGCTTACCATGCCAACGCAATAGCTAGAAAACGATCCACGGGTAA 3695
Qy 1041 CTCCAATTTGACATTTTCATAGCGAAACAAAGTTTAGACGAGTTTACACAGTTTATTAAT 1100
Db 3696 CGAAGATTGGCTCATTTTATAAGACGACACTCAAT--GATGAACCTGCCACAATTTATCAAT 3754
Qy 1101 GTTTTAATTTGGCGATATGAGTCTAGTTGGTACAGCTCCACCTACAGTTGATGAATTTGAA 1160
Db 3755 GTTCTTAAAGGGGATATGTCTATTAGTTGGTCCAAGACCAATCTGCTTTTGAAGCGAAA 3814
Qy 1161 AAATATATCTCTGCTCAAAAGAGAGAGATGAGTTTAAACCCAGGGAATTAACAGGTTCTTGG 1220
Db 3815 GAATATGGGAAACGCTTCGCTTACTTACTCATGTGCAACCCAGGAATCACTGTTATTGG 3874
Qy 1221 CAGGTTAGTGGTCTAGTAATATACAGACTTTCGACGACGTPAGTTTCGGTTGGACTTAGCA 1280
Db 3875 ACACACATGCTCGAAGTAAAGTTCTTTTCTCAACGAGCAGAT--TTAGAACTCTAT 3931
Qy 1281 TACATTGATTAATTGGACTATCTGCTCAGATATTAATAATTTTAAAGACAGATGAAAGTT 1340
Db 3932 TATCTCCAGTACCATAGCACCAAAATGATATCAAGCTTCTAGTACTACAAATTTGTACAA 3991
Qy 1341 GTATTGTTGAGAGAGGGAAGTAAAGTATATGAAAGTTTGTGTTGGTCCGTTCTTCA 1400
Db 3992 AGTATTAAACGATCGGACGATATAAATA--ATGAATAGCAATTAGTAGGTTCCAGC 4048
Qy 1401 GGGGACATTTGACTCTCTGTTGTTTAAACCCGTTTGGAAAGGAAGAAACGTTT 1460
Db 4049 GTGGCCATTTGACACACCTGTTTCTTAAATAAGTTTGGGAAAAACGAAGATAGATT 4108
Qy 1461 TGGTAAATTTGATAAAGAGGATGCAAGAGTCTTTTGAAGATGAAAAATGATATCCA 1520
Db 4109 TGGGTCAATTTGATAAAGCAGATGCAAAATCTATATTGAAAGAAAGAAATTTATCTCT 4168
Qy 1521 TGTTACTTTCCACAAATCGCAATCTCATTAATTTTAGTGAATAATCTTTCTTAGCTTTC 1580
Db 4169 TGTTATTTCCCAAAATAGAAATGTAAAAACACGATAAAAATACCAATTTCTTGCAATTT 4228
Qy 1581 AAAATTTTACGTGATGAGAAACACGATGTTTATTTATTTTATCTGTTGGCGGCTTGTCTC 1640
```

Db 4229 AAAATACCTTAGAAGAAACACAGATTGCTGATTTTCGAGTGGTGGTGGTAGCGGTT 4288  
Qy 1641 CCCTTCTTTTACATCGGAAACCTATTTTGGAGCAAGAGATTTATATTTGAAGTATTTGAT 1700  
Db 4289 CCTTTTCTTTGGTTAGGTAACATTTTCGGTGCAGAGACAGTCTATATTTGAATATTTGAC 4348  
Qy 1701 CGAGTTAATAATCTACATTAACCTGGAAGAACTAGTTTATCCCGTAAACAGATATTTTATT 1760  
Db 4349 CGGATGATTAACCAACCTTAAACAGGAAATAGTTTATCCAGTTACTGATAAGTTTATA 4408  
Qy 1761 GTTCAGTGGGAAGAAATCAAGAGGTATATCTTAAATCTATTAATCTTGGGAGTATTTT 1820  
Db 4409 GTTCAATGGGAGAGTTTAAAGAAAGTTTACCTTAAGCAATTAATTTAGGAGNAITTC 4468  
Qy 1821 TAATGATTTTGTAACTAGTGGAACTCATGAACCAACAGTTTAACTCGATGATAAAGAGA 1880  
Db 4469 TAATGATTTTGTAACTGGAACTCAGCAACCAACATTTAACTCGACTCATTCAGAAAA 4528  
Qy 1881 TTGATTTTGTAAAGAAATGGAAGTATAACCGAGCAATATTTTATTTTCAACAGGATTT 1940  
Db 4529 TTGATGAACCTGTACGGATGGTGAATCGAAGACGATGATTCATGCAAAATTTGGGTACT 4588  
Qy 1941 CTGACTATATTCAGAAATATTCAGATATATAAAGAAATTTCTCAGTTACAAAGAAATGGAAC 2000  
Db 4589 CAACCTATGAACCTAATATATCTAAATGGGAAAGTTTATTTGATGAGACTATGAAA 4648  
Qy 2001 AATATATTAACAATCAGAGTAGTTATTTGCCAGGAGCCCGCTACTTTTATGAAAT 2060  
Db 4649 GATGATGAATGAACGAGTACGATTTATCTCATGCGGACCATCTACCTATATGCAAG 4708  
Qy 2061 CATATCAAGGAAAGAAACAAATATTTGTTCTAGACAAAGAAAGTATGTTGACATG 2120  
Db 4709 TATTAACAAGTAAAGTAAATTTCCGATGTTGTTCCACGGCAATGAAATTTGATGAGCATA 4768  
Qy 2121 TAAATGATCATCAA-GTAGAGTTTGTAAAGAAATTTTCAAGATAATAATATTTTATTT 2179  
Db 4769 TAAATGATCATCACTTTGGTGAAGTAAACAGTTGTGAAAGAGGATCTCATTTGATTT 4828  
Qy 2180 --ATAGAAATATAGATGATTTGTTTCAAGAAATATTTGAAGTTTCTPAAGCAACTAAT 2237  
Db 4829 TGTGGAAGATGTTGAAGACATCTCGAAATATTTATAGTTCCAAATTTTCAGATACCT 4888  
Qy 2238 TTACATCAATATAATTTTGTGTAAGATTAACAAATAGTTGAAATTTAATG 2297  
Db 4889 TACAAAAAATGTAATCAACACACTGAATTCATATAAATTTATTCAGTGTGAAATTTACC 4948  
Qy 2298 AG 2299  
Db 4949 AG 4950

## RESULT 23

US-09-815-242-9587  
; Sequence 9587, Application US/09815242  
; Patent No. US2002061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815.242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9587  
; LENGTH: 1368  
; TYPE: DNA  
; ORGANISM: Streptococcus pneumoniae  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1368)  
US-09-815-242-9587

Query Match 6.1%; Score 423.2; DB 9; Length 1368;

Best Local Similarity 58.6%; Pred. No. 4.1e-50;

Matches 794; Conservative 0; Mismatches 548; Indels 12; Gaps 3;

Qy 14 ATTGGCAATTTTGTATATGATAGCAGTTGCAATTTCTGCAATCTTAAACAGTCAATATACC 73  
Db 27 ATTGGCAATTTTGTATATGATAGCAGTTTCTGTTATTTTATTTGACTTATCTACTTAGTCTGTGAG 86  
Qy 74 AAATGCTGATTTAAATCGTTCTGGAATTTTATCATTAATGATGTTTCATTTATTTTGCATT 133  
Db 87 AGAAGCGAGATTTGTTTCAACCAACACTATTTGCACTTTATATCTCTCATTTATTTTGTCTT 146  
Qy 134 TTTTATATCTCGTATGCCAGTTTGAATTTTCAGTATAGAGGTAATCTGATAGAGTTTGAATA 193  
Db 147 TTATATCAGTGAATATGAGCAGGATTTCTTAAAGAGAGATATTTGATGAACCTTGCCA 206  
Qy 194 AACATTTAATATATGATATATATTTTTCGAATTTTCTTACGCGAGTATCATTTTGTGGA 253  
Db 207 GACATTTGAATATATCTATTTCTTGCACACTAGCGATTTGTTTCTAAATTTTCTTTAGA 266  
Qy 254 GAATAATTTCCGACTTTTCAAGAGCTGCGCGTATTTTACATTAATTAACCTTCGTTT 313  
Db 267 GGATCGATTTAGTATTTCCAGACGAGGATGATTTTCTCTCACATTAATGCTCTCTT 326  
Qy 314 GGTATACCTATTTTAAACGTAATTTTAAAGCAGTTTAAAGGATGAGTCTTCTATTTTCGCAAT 373  
Db 327 AGTCTATGCTTAAACCTATTTATCAAGTGTATTTGGAAGCGGCTTATCCCACTTTAA 386  
Qy 374 CTATCAAAAAAGACGATTTCTAATTAACAGCGCTGAACGATGGAAGAAATATGCAAGTTT 433  
Db 387 AGGAAGTAAGAGATTTCTTACTTACAGCAACTTCTCGTGTGCAAAAGGCTACTGGATAG 446  
Qy 434 ATTTGATACATATAACAAATTTCAAAAAATCTTTTGCATTTGTTAGTGTAGTACAGAGA 493  
Db 447 ATTAATAGATCAAAATGAGGTTGTTGGGAAGTTTGGTAGCGCTCAGTGTCTTTAGATAACC 506  
Qy 494 AATAGATAAAATTAATTTTATCAATTCAGCTCTATTTTCTGTGGAAGAGCTATAGAGTT 553  
Db 507 AGATTTTCAGCATGATTTTAAAGGTAGT-----AGCAGAGGGGGAGATAGTAACTTT 560  
Qy 554 TTCAACAAGGGAAGTGGTGCACCACTTTTATAAAATCTTACCAAGTGAAGTTTTTAGACGT 613  
Db 561 TCGCACTCATGAGGTCGATGAAGTCTTTTATCAATCTTCCAGGTGAAAAATACATAT 620  
Qy 614 AAGCAATTCGTTTCAGATTTTTCAGTTTGTAGGTTTGTAGTATGTAAGCGTTGATATTAATTC 673  
Db 621 TGGAGAGCTTGTCTCTCAGTTTGAACGATGGGAATTTGATGATATAGTCAATCTTAAATGC 680  
Qy 674 ATTCGGTTTTTACTGCGTTGAAAAACAAAAATCCCAACTCTAGGTGACCATAGCATTTGT 733  
Db 681 TTTTGTATCGTAGTTTGGCACGTTAACCAAGCAAAATTCGTGAGATGCGAGGGTTAAACGTTGT 740



QY 734 AACTTTTCCACAAATTTTATAGCCCTAGTCATATCATGATGAACGACCTTTGGATAT 793  
Db GACTTTTTCTCAACAATTTTAAAGACTAGTCATGTAATTCGTAAGCGGATTTATGATAT 800  
QY 794 ACTCGAGCGTAGTCGGGTTAAATTTATTTGGTATAGTTCTATTTTGGTATTTCCAAT 853  
Db CGTGGTGCAATGGTAGGCTGATCTATGTTGGTTAGTCAGTATGTTGCTGTTCCCTT 860  
QY 854 TATTCGTAGATGGTGGACCGGCTATTTTCTCAGAAACAGTTGGACAGAAATGGACG 913  
Db GATTGAAAGGATGGGGATCTGCTATTTTGGCTCAGACGCGTATAGGAAATGGTCG 920  
QY 914 CATATTTACATCTCAAGATTTGATGATGATGATGATGATGATGATGATGATGATG 973  
Db TCAGTTCACTTTTATAGATTTGCTCTATGTTGATGATGATGATGATGATGATGATG 980  
QY 974 CTGCTCAGCAACACAGATGCAAGGTTGGTATGTTTAAATGGGAAACGATCCT 1033  
Db ACTCATGGAACAAATACCATCCAGGTGGATGTTT-----AGGTGGACGATGATCCT 1035  
QY 1034 AGAATTACTCCAAATGGACATTTTCATCGCAAAACAAAGTTTTCAGACGAGTTTACCACAGTT 1093  
Db CGTATCAGCAAAATTTGGTTGTTTATACG-GAAGACTAGCTTTGGACGAGCTTACCACAGTT 1094  
QY 1094 TTATAAGTTTAAATGGCGATATGAGTCTAGTTGGTACAGTCCACCTACAGTTGATGA 1153  
Db TTATAAGTTTCTAAGGGGAGATATGAGTTTGGTTGGCACACGGCCACCAACAGTGGACGA 1154  
QY 1154 ATTTGMAAAATATCTCTGCTCAAAAGAGACGATGAGTTTAAACCGGGATTTACAGG 1213  
Db GTATGAGCACTATACCCAGAAACAAACGGCGACTAAGTTTAAACCTGCGATAACAGG 1214  
QY 1214 TCTCTGGCAGGTAGTGGTCTAGTAAATATACAGACTTTCAGACGAGTAGTTCCGGTTGGA 1273  
Db TTTATGACAGTTCAGCGGACGAGTGAATCAAGATTCAGATGAGTTGTCAAATAGA 1274  
QY 1274 CTTAGCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1333  
Db TGTGGCTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1334  
QY 1334 GAAAGTTGATGTTGAGAGGAGTAAAGTAA 1367  
Db TAAAGTAGTATTTATGAGAGATGAGCGAAGTAA 1368

## RESULT 24

US-10-804-408-178  
; Sequence 178, Application US/10804408  
; Publication No. US20040253617A1  
; GENERAL INFORMATION:  
; APPLICANT: FANRONG, KONG  
; APPLICANT: GILBERT, GWENDOLYN  
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI  
; FILE REFERENCE: 675002-2001  
; CURRENT APPLICATION NUMBER: US/10/804,408  
; CURRENT FILING DATE: 2004-03-19  
; PRIOR APPLICATION NUMBER: PCT/AU02/01281  
; PRIOR FILING DATE: 2002-09-18  
; PRIOR APPLICATION NUMBER: AU PR 7749  
; PRIOR FILING DATE: 2001-09-19  
; NUMBER OF SEQ ID NOS: 182  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 178  
; LENGTH: 2581  
; TYPE: DNA  
; ORGANISM: Streptococcus agalactiae  
US-10-804-408-178

Query Match 6.0%; Score 419.8; DB 18; Length 2581;  
Best Local Similarity 51.9%; Pred.No.1.6e-49;  
Matches 1100; Conservative 0; Mismatches 972; Indels 47; Gaps 5;  
QY 3063 GAATATTTTATTTTAAAGTACTTCTGGTTGATTTTATTTTCCAGAGCAAAAGTATGTA 3122

Db 492 GATTATTTCTTTATTTATGCTCTTTGGGTACTTATTTTAGTACCAAACTATGTTATCAG 551  
QY 3123 TTTTATTTAAATTTTATGAATTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTT 3182  
Db 552 TTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 611  
QY 3183 ATATTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 3242  
Db 612 ATAGCAATTTCTCAATACTATTTCTGCTTTGGTTATTTATTTATTTATTTATTTTATTC 671  
QY 3243 GTCAAGATGTTTGTGAAATTTTGAAGATTTTGAAGATTTTGAAGATTTTGAAGATTTT 3302  
Db 672 ATTAGAGTACTCAAGAGGATATACGTTTTCAGCGATTTTATTTGCTGAGCTATTTAAACT 731  
QY 3303 ATAAATTTGGATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGA 3362  
Db 732 ATTAGTACAGATATGCTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 791  
QY 3363 AAAAAATTTAAATATGATCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 3422  
Db 792 TCAGTTGTAAGGAATGTGTAAGGTTAACTATTTTGTGTTGTTTCTTTATTAACAGTTT 851  
QY 3423 TATATTTTCAAATGGGAAGATTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTT 3482  
Db 852 TATTTATTTTCTTATGCTGAGCCAACTTTTATTTGGAAGAGAAATGTTTTCATATAG 911  
QY 3483 TATCTTATAACAGCGCTCAAAACAAAGTTGGTGGCTTTTATGAACTATCTTACGTTTAA 3542  
Db 912 TGGTTT---CCATATATGAGATTAAGACTTGGCGATATTTTGAATATGCTTACACTAAT 968  
QY 3543 ACCACTCAATATATGTTTCAAATTCGTTAATCTTTGCACTTATATAAAATAAATAATG 3602  
Db 969 GGTCAAGTTTATTTTATTTTCTTATCCCATACTTTTGTGAAACCCCAAAACATATGAA 1028  
QY 3603 CAATTTTCTTGTGCTGCTTTTATACCGATCTATTTAAGTGGATCGAGAAATGCT 3662  
Db 1029 AATATTTTAAATCTTACTGTTGACTATATGTTTCTTCTGCGCTAGAACTATA 1088  
QY 3663 AGTTTATCGCTAGCAATATTAATTTATGCTTGTATGAGATATATAGTTGGAAATTT 3722  
Db 1089 TTGCTCTGATGTTGGTTTATTTAGATCGCTTCTTTTATTTATTTTATTTTATTTTAA 1148  
QY 3723 GCTTGGATAAAAAAGCTAATAGTAAATTTGTAATCTACTTATTTTATTTTAAATA- 3777  
Db 1149 AATTTGAAATTTGACCAAGAAAACACATTTTATATACTTGGTATGACTTCTTATTTAT 1208  
QY 3778 -----CTGAATTTGCTTTACCATGAAATTTTGGCTGTTTATTAATCTTAGAATCA 3827  
Db 1209 GCTTGTTTTCTTATAACATATGCTCAATATTTGAAAAAAATAATTTATGTCAGAAAC 1268  
QY 3828 AGTAAACGAGCTAGATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 3887  
Db 1269 AGTACTATCACTAGGATGATGTTTATCAAGAAAGTATTTTGAAGTTTCTAAAAAGGA 1328  
QY 3888 ATTTTATTTGATATGAAATATCCGAATATTCAGTTACGGAACTTGGCTCGGAAGTCAT 3947  
Db 1329 ATTTTATTTGACAGGTTATAGGATTTCCATCAAGTGAAGAAATTTCTTAGGATCGAT 1388  
QY 3948 TCAGGCTATATATCAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 4007  
Db 1389 TCTACTTATATAGTGTCTTTTACAGGACTTCTTTATTTAGGAAATGTTTCTTTATTTTCT 1448  
QY 4008 TCITTTTTTATGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 4067  
Db 1449 GCCTTTTATCTTTTATATAAAGAGCGGATTTCAAAAAATTTATTAATCTCAGATATTT 1508  
QY 4068 TTTACATCAATTTAGCCATATTTTTCATATATGAAACAATAGATCCGATTTATTTATTTAT 4127  
Db 1509 TTTTATAGGTTATTTATGTTACACGCTTTTGAGAAATAGATCTTAATCATTTGAGATTT 1568  
QY 4128 GTACTATTTCTTTCTTCAATAGGTTATTTGGAAATATAATAATTTTAAAAAGGATATGG 4187



Db 1569 GTATTATTATTCACACTTTTGGTAT-----AGTGGAGGG 1605  
 Qy 4188 ACAAAAATGAATGATTTAATTCAGTTATGTACCAATTTAATGTCCTCAAGATTAATCT 4247  
 Db 1606 CTAAGAAATGAAGAAAGTAAACAGTCATTTATACCTATATACCACTCAGAAGCATACCT 1665  
 Qy 4248 TGATAAATGATTAACAGTATTTAACCACCAATATACCTAATTTAGAGGTTATTCCTCGT 4307  
 Db 1666 TAAAGATGTGCGAATCCGCTACTACACAGACTCATCCATTTGATGAAGTTTATCTAAT 1725  
 Qy 4308 AAATGATGGAAGTACTGATGATTCAGAAAATTTGCTTAAACTATATGAAGAACGATGG 4367  
 Db 1726 TGATGATGATCCACTGATTAATAGTGGAGAAATTTGATGAATTTATCTCAGGAAGTAA 1785  
 Qy 4368 AAGAAATTAATATTAACAAGAAAATTAATGCGGCTAGCAGATGCTCGAAATTTCCGACT 4427  
 Db 1786 TCGCATACTTGTATTTTCATAAAAAATGAGGGGCTCTTCGCGAAGGAACCTAGGTCT 1845  
 Qy 4428 AGACATGCAAGTAAATATATTTGCTTTGTCGATTCCTGATGACTATATAGAGTTGC 4487  
 Db 1846 AGATAAATCCACAGGAGAAATTCATAACATTTGTGGATAGTATGATTTTGTAGCCGAA 1905  
 Qy 4488 AATCTCCAGAGATGCATGATATATAAATGAGTATAATGCGGATATAGCAGAGATAGA 4547  
 Db 1906 TATGATGAATATGTTAAAAATTTAATCACTGAGATGCTGATATAGCAGAGTAGA 1965  
 Qy 4548 TTTTGTGTTAGTAGCGAAACCGGATATACAAGAAAAAAGAAATAGTAAATTTTCATGT 4607  
 Db 1966 TTTTGATATTTCCGAATGAGAGAGATTTATAGAAAGAAAGACGAACTTTTATAAAGT 2025  
 Qy 4608 CTTAACGAGAGAGACTGTAAGAAATTTTGTGAGGATCTAATATAGAAAAATATGT 4667  
 Db 2026 TTTTAAAGAAATTAACCTCTTTGAAAGAAATTTTATCAGGTAATAGTGGAAAAATATGT 2085  
 Qy 4668 TTGTGCAAGCTTTATTCACGAGATATTAATAAGATATAAAATTCCAAATTAATATAG 4727  
 Db 2086 TTGTACAAATATATAAAAAAGTAAATTTGTAACCTTGAGGTTGA---TGGAACCTT 2142  
 Qy 4728 AAGTATTTGGTGGAGATTTGCTTTTAAATTTGGAGGCTCTGAACAAATGTAACAGTGTAGT 4787  
 Db 2143 AAAAAATTTGGTGGAGATTTACTTTTAAATTTGCAACTCTTATGTCAAGAGCACCGTATAGT 2202  
 Qy 4788 AGTTGATCTAGAGAAATATTTATTAATATATGTCATTCGTAACAGTTTCGCTTATTAATCA 4847  
 Db 2203 CGTAGATACGACTTCTTCCCTTATATACTTATPCGAATTTGTAAAACTTCTGTAAATGAATCA 2262  
 Qy 4848 GAAATTCCTATAAATAATTTGATTTAGTCACAAAGATTCGAGAAATTAACCCCTTTAAGTT 4907  
 Db 2263 GAAATTCACGAAACTCATTAGATTTTATAACAAATTTTAAATGAATAGTAGTTGGT 2322  
 Qy 4908 AAAAGAGAGTTTATGTCATTTATTTGATGCAAAAGTTTATTAAGAGAGAGTTTAAATGTTTT 4967  
 Db 2323 TCCTGCCAGATTTAGCTAATTTATGTTGAAGCGAAATTTTAAAGAGAAAGATAAAGTGTCT 2382  
 Qy 4968 AACAAATGATTTCAACAGATTTG---TTTGGTAATGAGTTCTTGGCAATATAGAGTC 5024  
 Db 2383 CCGAAAAATGTTGAAATTTAGGTAGTAATATTTGAACATAAATCAAAGTACAACGAGAGAT 2442  
 Qy 5025 TTATCGAAAAAGAAATACGTAGATATCCATTTATTAAGCGAAAGATATTTATCAAGAAA 5084  
 Db 2443 TTTTTCACAGACATTAATATACACCGTCTTATTAAGCCGTCAATATCTTATCATTA 2502  
 Qy 5085 GCATTTAGTTACGTTGATTTGATGAAATTTTCCCTTAACATAATGATGTAATGTTATATA 5144  
 Db 2503 GGGATTAATGAAGCTTTTATTAATGAATGTTCCACCTAACTATATGTTGTCATATAG 2562  
 Qy 5145 GAAATTTCAAAGCAGTAG 5163  
 Db 2563 AAGATTTCAAAGCAGTAG 2581

RESULT 25

US-10-804-408-179

; Sequence 179, Application US/10804408  
 ; Publication No. US20040253617A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FANRONG, KONG  
 ; APPLICANT: GILBERT, GWENDOLYN  
 ; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI  
 ; FILE REFERENCE: 675002-2001  
 ; CURRENT APPLICATION NUMBER: US/10/804,408  
 ; PRIOR FILING DATE: 2004-03-19  
 ; PRIOR APPLICATION NUMBER: PCT/AU02/01281  
 ; PRIOR FILING DATE: 2002-09-18  
 ; PRIOR APPLICATION NUMBER: AU PR 7749  
 ; PRIOR FILING DATE: 2001-09-19  
 ; NUMBER OF SEQ ID NOS: 182  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 179  
 ; LENGTH: 2577  
 ; TYPE: DNA  
 ; ORGANISM: Streptococcus agalactiae  
 ; US-10-804-408-179  
  
 Query Match 5.5%; Score 386; DB 18; Length 2577;  
 Best Local Similarity 51.6%; Pred. No. 8.9e-45;  
 Matches 1092; Conservative 0; Mismatches 975; Indels 51; Gaps 7;  
  
 Qy 3063 GAATATTTATTTTAAAGTACTCTCGGTGATTAATTTTATCCAGAGCAAAAGTAGTA 3122  
 Db 492 GATTATTTATTTTATGCTCTTTGGTACTTATTTTAGTACCAACCAATGGTATCAG 551  
 Qy 3123 TTTTATTAATTTTATGATTAATTTTATTTTATCATATAAAATTTTGAACACTAAGCTA 3182  
 Db 552 TTTTAAATTTTACCATTATAGTCTTATTATTACTTTTGAAGAGTGAGTTTAGAATATCT 611  
 Qy 3183 ATATATAAAATGAAATTTTATTTTATTTATGCTCTATATTTATTTTGTTCAGTA 3242  
 Db 612 ATAGCAATCTTCAATACTATTTCTGCTTGGTATTTATTTATTTTGGCAATCTC 671  
 Qy 3243 GTCACAGTATGTTGTTGAAATAAATTTGAAAGATATTTGACAGATTTTACTGCTCC 3302  
 Db 672 ATTAGAGGTACTCAAGAGGATATAACGTTTCAGCGATTTATTTGCTGAGCTATTA 731  
 Qy 3303 ATAAATTTGGATTTATGCAATAATGATTAATAATTTGATTTCAATTTAATAATTTGATTAT 3362  
 Db 732 ATTAGTACAGGATATGCTTTATTTTATATAATTTATAGAAAAGCTGATTTTAATAGT 791  
 Qy 3363 AAAAAATTAATAATAGTATCTTTTATGTTTTTATGTTTTTATGATATATCTGCAATG 3422  
 Db 792 TCAGTTGTAAGGAATGTTGTAAGGTTAACTATTTTGTGTTGTTCTTATAACAGTTTTA 851  
 Qy 3423 TATATTTTCAAAATGGGAAAGATATGATTTTATTTTATAGACAGACACCTTATAGGACTAGAC 3482  
 Db 852 TATTTATTTTCCAAATGAAATTTACTACATTTCTAGGAAGAGATTTATTTTCAATTTGAA 911  
 Qy 3483 TATCTTTAAACAGGCGCTCAAAACAAGTTGGTGGCTTTTATGAACTATCTCTAGCTTAAAT 3542  
 Db 912 T---GGATTCCTTCTATGAAGTTAGACTTACTGCAATTTTGGATATGCAACTATTA 968  
 Qy 3543 ACCACTACAAATTAGTGTTCATTTCCGTTAACTTTTGCCTTATAAAAAATAAAATGCAA 3602  
 Db 969 GGTCACTTTATTTTATTTCACTATCCGATATATTTTAAACAGCAGAGGTATGGAGAA 1028  
 Qy 3603 CAATTTTCTTGTGCTGCTTTTATACCGATCTTTTAAAGTGGATCGAGAAATGGT 3662  
 Db 1029 AATATTTTATCACACTATTTCTTAGTTTTTGTGCATATTTTGACAGGGGCAAGAAATTTTC 1088  
 Qy 3663 AGTTTATCGTAGCAATATTAATTTATGCTGTTATGAGATATATAGTGGAAATTTT 3722  
 Db 1089 CTAAATTTGATGATAATTTTATAGGTATTTTACTCTTGAATAATCATTAATAAATTT 1148  
 Qy 3723 GCTTGGATAAAAAGCTAATA---GTAATATTTGTAATACTACTTATTTTAAATFAC 3779  
 Db 1149 AACCTAAAAATTAATAAAAAAGCTGCTTTTTCGATAATATATGGGATAAATATTATTATTG 1208

DB	2259	ACAAATTAATAAAATAACATTAGACTTTCATTGTATATTTTAAATGAGATTCATCAGGATAG	2318
QY	4908	AAAAAGAGAGTGTAGTCATTATTTTGGATGCCAAAGTTATTAAAGAGAAAGGTAAATGCTTT	4967
DB	2319	TCCGACAGAAATGTTTAATTTATGTGGAAGCGAAGTTGTACGAGAAAATAATCAAGTGT	2378
QY	4968	AAACAAATGTATTCACACAGATTTGTTGGATAATAGTCTTCGCCAATA---TTAGAGTC	5024
DB	2379	AAGGAAATGTTTGAATTTAGGAGAAATAGCTGATGAAATTTACGTTTACAGAGATATAA	2438
QY	5025	TTATCGAAAGAATACTGTAGATATCCATTATTTAAAGCGAAAGATATTTTCAAGAAA	5084
DB	2439	ATTTTGGCAGAGATATTAAATCATATTCAAATATGCAAGCAATAGGTTCTTATCTAA	2498
QY	5085	GCATTAGTTACGTTGTATTTGGATGAAATTTTCGCCCTAAACTATATGTAATGTATATAA	5144
DB	2499	ACATATCTGTACGTTATTTTGGATGAAATATTTTCGTACGTATATAAAGATGTATAA	2558
QY	5145	GAATTTCAAAAGCAGTA	5162
DB	2559	TAAATTTCAAAAGCAATA	2576

Search completed: April 29, 2005, 01:43:00  
Job time : 3641 secs

**This Page Blank (uspto)**



```
Qy 1911 CCGACGAAATATTATTCAAAACAGGATATTCTGACTATATTCAGAAATATTGCAAGTATA 1970
Db 1825 ATAAAAAAATNTAATAAAAAATAAATAATATAATATAAAAAATAAAAAATAAAAAA 1766
Qy 1971 AAAAAATTTCTCAGTTACAAAGAAATGAAACAATATATTAACAAATCAGAAAGTAGTTATTT 2030
Db 1765 ATTAAATATATAAATAATANAATTTAAATTAATAATATAATAATATATTAAATATTAAT 1706
Qy 2031 GCCACGAGGCCCGCTACTTTTATGAATTCATTATCCAAAGGAAAAAACAATTAATGTT 2090
Db 1705 ATTNATATAAATAAAAAATNTAAATTTATATANAATAATAATTTAAAAATAAATAAATA 1646
Qy 2091 TTCTTAGACAAAAAGTAGTGGCAACATGTAATGATCATCAAGTAGAGTTTGTAGAA 2150
Db 1645 TAATTANNAAAAAATAAATTTATTTTNAATTTATATATAATAATAATTAATTTTAAATA 1586
Qy 2151 GAATTTTACAAGATAAATAATTTTATTTATAG-----AAAAATATAGATGTTTG 2200
Db 1585 TNATTATAAATATATATATATTTTATTTTATTTTATTTTATTTTATATATAATATTTT 1526
Qy 2201 TTTGAAAAAATTAATGAAGTTTCTTAAGCAAACTAACTTTTACATCAATAATAATTTTTT 2260
Db 1525 TAATAATTTTATTTAAAAATAATAATATANAATAAAATATATTTTATTTTATTTTATTT 1466
Qy 2261 TGTGAAGAGATAAACAATAAGTTGAAAAATTTTAATGAGGATCAAGAAATGAATAATA 2320
Db 1465 TTTAAAAATATTTATATTTATAAATAAAATTTAATAAATAAAAAATAAATAAATAATTT 1406
Qy 2321 AAAAGATGCAATATTTGATAATGGCTTATCATATTTTCTCAGATTTTACTGGAGAGGA 2380
Db 1405 ATTTTATTAATATTTTAAATAATATTTTAAATTTNANNNAATTAATAATATAATAAAT 1346
Qy 2381 TACAGATATTTATCATCTTCTCCAGGAGAAATGCACACCATAGTTCTTCAGAAATACCTG 2440
Db 1345 NATAATTTAAAAATAATTAATAATTTATTTTAAATTTTAAATTTTAAATTTTAAAAATA 1286
Qy 2441 TATAATTTATTTTAAATATCTCAGGATTTATATGTTGAATTTTACAAGATGACAAAAA 2500
Db 1285 AAAAAAATAAATTTATATAAATAAATAATATATNAAAAAATAAATAAATAAATAAATAA 1226
Qy 2501 TATAAGAAAAATAGGATATATGAACGAGTTAAATGTTTACAGATTTATTTCTTAATATCA 2560
Db 1225 TATATAATTAATATTTATATAAATAAATAAATAATTTTAAATAAATAATTTTAAATAT 1166
Qy 2561 GAAAAAATAATGATAATGTAAGTCTTTTGAATTTTATTAAGAAATGTAACGAGCTTTTGAA 2620
Db 1165 AATAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1106
Qy 2621 TACTATTACAAAGATTTGTTGTTTATGATAGAAATAAATAAATAAATAAATAAATAA 2680
Db 1105 AAATNTATTTTATTTTAAATTTAATAAATAAATAAATAAATAAATAAATAAATAAATA 1046
Qy 2681 TTGGTTCTAAATGGGTTTCGCTCCACATGATTTTGGGCAATTTCTTTTATCAAAATGAA 2740
Db 1045 TATTTNTTATATATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 986
Qy 2741 ACGAACACAGCTTATTTTATTAAGTAATCTAAATGTCAGAGTGAACACTTTTATACAGCAA 2800
Db 985 --TANATATTTTATTTTATTTTATTAATAAATAAATAAATAAATAAATAAATAAATA 928
Qy 2801 TTATAGAAAAATATGAATTTTCAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2859
Db 927 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 868
Qy 2860 AAGTGGAAAAAATCAACATCTTCTCCTTATGCTTTTACAGATGATCTTATTTGATGATTC 2919
Db 867 AAAAAATANAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 808
Qy 2920 CTAATGCAAGAAATTTAGGTTTTTTTATTTGCTAGAAAGTTAAAAATAGAAAAATAAATCT 2979
Db 807 TAAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 748
```

```
Qy 2980 ABAATTAAGAAATATTATTACTAAAAAATAAATAAGTAGTTGATTTTGTGAGAGTAATGATGT 3039
Db 747 ABAATTAAGAAATTTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 688
Qy 3040 TTAATAATTTTAAAA-----TATGACCCGGAATATTTTATTTTAAAGTACTCTCTGGTTGAT 3094
Db 687 TTTTATTTTATTTATTTTATTAATAATAATAATAATAATAATAATAATAATAATAATAATA 628
Qy 3095 TATTTTATTTCCAGACAAAGATGATATTTTATTTTAAATTTTATGAAATTTTAAATTTTAT 3154
Db 627 AATTATNATTTATTAATAATATTATATAAATAAATAAATAAATAAATAAATAAATAAATAA 568
Qy 3155 TCAT--ATAAATTTTGAATACTAGCTAAATTTAAAAAATGAAATTTTATTTGTTTTTA 3212
Db 567 TNATAAATAAATAATATTATATATANTTAATAATAATAATAATAATAAATAAATAAATAA 508
Qy 3213 TTATGCTCTATATTTATGTTTGTTCAGTAGTCACAAGTAGTTGTTGTTGAAAAATAATTTT 3272
Db 507 TTATTAATAATTTTATATANAATTTATTAATTTATTTATTTATTAATAATAATTTTATTA 448
Qy 3273 GAAAGATTTATTCAGATTTTACTGCTCCCATTAATTTGGATTTTGCATAATGCAATAATGAT 3332
Db 447 TTTATAAATAATAATAATAATAATAATTTTAAATTTTAAATAAATAAATAAATAAATAA 388
Qy 3333 AATTGCTATTTTATTAATAATTTGATTTATTAATAAATAAATAAATAAATAAATAAATAA 3392
Db 387 TTATTAATAATTTTATTAATAATTTTATTAATAAATAAATAAATAAATAAATAAATAA 328
Qy 3393 TTTTATGTTTATTTAGGTATATCTGCATTTGATATTTATTTTCAAAATGGAAGATATTGTA 3452
Db 327 TAATAATAATAATAATAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 268
Qy 3453 TTTTATGACAGACACCTTAT 3472
Db 267 TAATTAATAATAAATAATAT 248
```

## RESULT 2

CG753083

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

CG753083 1896 bp DNA linear GSS 24-OCT-2003  
P048-1-C01.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,  
genomic survey sequence.

CG753083 GI:37977199

GSS.

Pristionchus pacificus

Pristionchus pacificus

Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;

Neodiplogasteridae; Pristionchus.

1 (bases 1 to 1896)

Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,  
Buntjer,J., van der Meulen,M. and Sommer,R.J.

An integrated physical and genetic map of the nematode Pristionchus  
pacificus

Mol. Genet. Genomics 269 (5), 715-722 (2003)

22835951

12884007

Contact: Sommer RJ

Evolutionary Biology

Max-Planck-Institute for Developmental Biology

Spemannstr. 37-39, Tuebingen D-72076, Germany

Tel: 00497071601371

Fax: 00497071601498

Email: ralf.sommer@tuebingen.mpg.de

Class: BAC ends.

Location/Qualifiers

1..1896

/organism="Pristionchus pacificus"

/mol\_type="genomic DNA"

/strain="California"

/db\_xref="taxon:54126"

/clone\_lib="Ppa EcoRI BAC Library"

/note="The library was generated by a partial digest of

the genomic DNA with EcoRI and cloning into the BAC vector."

## ORIGIN

```

Query Match      2.48; Score 167.4; DB 9; Length 1896;
Best Local Similarity 46.8%; Pred. No. 1.1e-15;
Matches 761; Conservative 0; Mismatches 830; Indels 35; Gaps 9;

Qy 2050 TTTTATGAATTCATTCCTCAAGGAAAGAAACCAATTTATGTTCTCTAGACAAAAGTA 2109
Db 219 TNATATTTATTTATTAATTAAGAAATTAATATATTTATATTTATTAATTAATTAAGAAANT 278
Qy 2110 TGGTGAACATGTAATGATCATCAAGTAGAGTTTGTGAAGAGAAATTTTACAGATATAAA 2169
Db 279 TANATATTANANATATTTANNANAAATTTATTTATTTATTTATTTATTTATTTATTTATAA 338
Qy 2170 TATTTTATTTATAGAAATATAGATGATTTGTTTGGAAAGAAAT-TAATGGAAGTTTCTAAGC 2228
Db 339 TATTTATTTTATAGAAATATATTTANNTATATTAATAAATATAAATTAATTTATTTATAA 398
Qy 2229 AAACCTAACTTTACATCAAAATAATATTTTCTGGAAGATTAAAAACAATAGTTGAAA 2288
Db 399 TTATATAATNATTTTAATAATTAATAATATATATTTATTTATTAAGAAATATAATA 458
Qy 2289 AATTAAATGAGATCAAGAAATCAATTAATAAAGATGATATTTGATAAGCTTAT 2348
Db 459 AAAATATTTAATAATAATAATTAATAATTAATAATAATTTAATAAATAATTTATAT 518
Qy 2349 CATTAATTTTCTCAGATTTTACGGAGAGGATACAGATATTTATCATCTTCTCAGGAG 2408
Db 519 AATTTTATATATTTATATATTAATAATATATATATTTATTTATTAATAATTTAAATTA 578
Qy 2409 AATGCACACCATTAAGTTTCTTCAAGATACCTGTTAATAATTTTAAATATTTCTCAGGAT 2468
Db 579 ATTATATAAANNAATTTATTTATATAATATTTTAAATATAATTAATTTATTAATTAAT 638
Qy 2469 TATATGTTGAATTTACAAAGATGAGCAAAATAATAAAGAAATPAGGATATAGAACGAG 2528
Db 639 ATATTTNNATATTTATTTATATTTATTAATAATTAATAAATAAATAATATATAT 698
Qy 2529 TTAATGTTACAGATTTATTTCTTCAATATATACAGAAATACTATTGA--TAATGTTACTGTT 2586
Db 699 ATATTTTAAATTTTATTTTAAATTTTATTTATTAATAATTTATTAATTTATTTATTA 758
Qy 2587 TAGAATTTTATTAAGATGATCCAGCTTTTGAATCTATTTTACAAAGATTTGTTGTTTAT 2646
Db 759 TAAATTAATATATTTAAATTAATAAATTTATTTATTTATTTTAAATTTTAAATTTTNA 818
Qy 2647 TGATAGAATAAAAAACATGGTCTAAGAAATAGATTTGGTTCTAATGGGTTTCGCTTCCA 2706
Db 819 TNATAAATTTATATAAATAATAATAATTTATTTATTTATTTATTTATTTATTTTAAATTT 878
Qy 2707 CATGATTTTGGCAATCTTTTATCAAAATGAA--AACGAACACGCTTATTTATTTAAGTA 2765
Db 879 ATTTATTTTATTTATTAATAATAATTTTAAATAATTTTAAATTTTAAATTTTAAATA 938
Qy 2766 ATCTAATATGCCAGATG-----ACTATTTATACAGACAAATTAAGAAATTAAGATTT 2820
Db 939 AATTTATTTTAAATATATTTTAAATAATTTAATAAATAAATAATNTATTTATTTATTT 998
Qy 2821 TCAATAGATTTATCTAAATATGAAATTTAAGATATATAAAGTGG----- 2865
Db 999 AAAATTTTATTAATAATTTTATTTATATAATAATAATAAATAATAATAATAATTTAT 1058
Qy 2866 ---AAAAAATCAACATCTTCTCTTATGCTTTTACAGATGATTTCTATGATGAATTTGCTA 2922
Db 1059 ATAAATAAATAAATAATTTATTAATAATTTAATAAATAAATAAATTTTATTTTAAAT 1118
Qy 2923 AATCAAGAATTTTGGTTT---TTTATTTCTAGAAAGTTTAAATAAGAAATAAATCT 2979
Db 1119 TATAATATTTATTTTATTTATTTATTTTATTTATTTATTTATTTATTTATTTATAAAT 1178
Qy 2980 AAATTTAAAGAAATTTATTTACTAAAAATAAATAAGTTGATTT---TGTGAGAGTAAGTA 3036

```

```

Db 1179 AATTTTATTAATAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1238
Qy 3037 TGTTTAAATTTATTTAAATATGACCCGGAATATTTTATTTTAAAGTACTCTCGTGTGATTA 3096
Db 1239 TATATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTAAT 1298
Qy 3097 TTTTATTTCCAGAGCAAAAGTATGTTATTTTATTTTAAATTTTATGAAATTTAAATTTATTC 3156
Db 1299 TTAATTTAAATAATTTAAAAAATAATATTTAAATTTATTTAAATTTATTTATTTATTTATA 1358
Qy 3157 ATATAAATTTTTCGAAACCTAAGCTAATATTAAGAAATGAATTTTATTTGTTTTATATAT 3216
Db 1359 TATTTATATTTNNNTAAATTTTAAATATATTTTAAATATTTTAAATAAATAAATAATTTANT 1418
Qy 3217 GGTCTATATTTATGTTTGTTCAGTAGTCAAGATGTTTGTGTTGTTGTTGTTGTTGTTGAAA 3276
Db 1419 AATTTTATTTT-TATTTNATATTTTATTTATTAATAATAATAATAATAATAAATAATA 1477
Qy 3277 GATTTATTTGAGATTTTACTGCTCCCATATTTTGGATTTATGCAATTAATGTTATTTAAT 3336
Db 1478 AAAAAAATAATAATTTTATTTATTTATTTTAAATAAATAATAATAATAATAATAATA 1537
Qy 3337 TGTATTTCAATTTATTAATTTGATTTAATAAATAATAAATAATAATAATAATAATAATA 3396
Db 1538 ATATTTAAAAATAAATAATAAATAATAATAATAATAATAATAATAATAATAATAATAATA 1597
Qy 3397 TAGTTTTATTTAGGTATATCTGCAATTTGTTATTTTCAAAATGCGAAAGATATTTCTATTT 3456
Db 1598 TTATTTATTTATTTATTTAATAATTTNAAATTAATAATAATAATAATAATAATAATAATA 1656
Qy 3457 TAGACAGACACCTTTATAGGACTAGACTATCTTATAACAGGCGTCAAAACAGGTTGGTTG 3516
Db 1657 TTTTAAATTTATTTATTTATTTATTTTAAATTTTATTTATTTATTTATTTATTTATTTAT 1716
Qy 3517 GCTTTATGAATCTACCTACGTTAAATACCACTACAAATTTAGTTTCAATTCGGTTAATCT 3576
Db 1717 TAATAATAATTTATTTATTTTAAATTTTAAATTTTATTTATTTATTTATTTATTTATTT 1776
Qy 3577 TTGCACCTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3636
Db 1777 TAAATTTTATTTATTTATTTATTTATTTTATTTTATTTATTTATTTATTTATTTATTTAA 1836
Qy 3637 TCTATT 3642
Db 1837 TTTATT 1842

```

## RESULT 3

```

CG750135/c
LOCUS      CG750135
DEFINITION P044-3-D09.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
            1592 bp DNA linear GSS 24-OCT-2003
ACCESSION  CG750135
VERSION     CG750135.1
KEYWORDS   GSS
SOURCE     Pristionchus pacificus
ORGANISM   Pristionchus pacificus
            Nematoda; Chromadorea; Diplogasterida;
            Rukaryota; Metazoa; Nematoda; Pristionchus.
REFERENCE  1 (bases 1 to 1592)
AUTHORS   Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
            Buntjer,J., van der Meulen,M. and Sommer,R.J.
TITLE     An integrated physical and genetic map of the nematode Pristionchus
            pacificus
JOURNAL   Mol. Genet. Genomics 269 (5), 715-722 (2003)
MEDLINE   22835951
PUBMED    12884007
COMMENT   Contact: Sommer RJ
            Evolutionary Biology
            Max-Planck-Institute for Developmental Biology
            Spemannstr. 37-39, Tuebingen D-72076, Germany
            Tel: 00497071601371

```



/organism="Pristionchus pacificus"

/mol\_type="genomic DNA"

/strain="California"

/db\_xref="taxon:54126"

/clone\_lib="Ppa EcoRI BAC Library"

/note="The library was generated by a partial digest of the genomic DNA with EcoRI and cloning into the BAC vector."

## ORIGIN

```
Query Match          2.3%; Score 158; DB 9; Length 1348;
Best Local Similarity 47.1%; Pred. No. 3e-14;
Matches 575; Conservative 0; Mismatches 631; Indels 15; Gaps 3;

Qy 2928 AAGAAATTAGGTTTTTTTATTTGCTAGAAAGCTTAAAGATAGAAATAAATCTAAATTTAA 2987
Db 128 AAAAAATTATTTTATNTNTTTTATTTATTTNATNAATATTTAAATATTTAAATTTAA 187

Qy 2988 AGAAATTATTAATAAAATAGTTGATTTTGTGAGAGTAAATGTAATGTTAAATTA 3047
Db 188 TATTTTTTTTTTATAAAATATTTTATTAATTAATAAAATTTTTTNTNTTTTATAATTTT 247

Qy 3048 TTTAAATATGACCGGAAATATTTTATTTTAAAGTACTTCTGTTGATTTATTTTATTTCCA 3107
Db 248 TTTATAAATTTTTTATTAATATTTTTTTTTTTTTTTTTTTTTTATTTATTTATTTT 307

Qy 3108 GAGCAA---AAGTATGATTTTTTATTAATTTTTTATGAATTTAAATTTTTTATTCATAAAA 3164
Db 308 TTTTATTTATTAATTTAAATTTTATTAATTTTTTTTTTTTTTTTTTTTTTTTTTTTATTT 367

Qy 3165 TTTTGGAAACTAGCTAAATTAATAAATGAAATTTTATGTTTTTATTAATGCTCTATA 3224
Db 368 TTTTTTTTTTTTAAATATTAATATATTTATTTATTTTTTTTTTTTTTTTTTTTTTTTTTT 427

Qy 3225 TTATGTTTTGTTTCAGTAGTCACAGTATGTTGTTGAATAAATTTTGAAGATTATTT 3284
Db 428 TAATTTTTTTTTTAAATTTATTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTATTTT 487

Qy 3285 GCAGATTTTACTGCTCCCATATTTGCAATTTGCAATAATGTAATTAATTTGTAATCA 3344
Db 488 TTTTATTTTATTTTTTTTATATTTATTTATTTTATTTATTTATTTATTTATTTT 547

Qy 3345 TTTATAAATATGATATATAAAATATAAATAGTATCTTTTTTTAGTTTTTTAGTTTTTA 3404
Db 548 TATTANATAATTTTTTATATTTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTATTTATTT 607

Qy 3405 TTAGGTATATCTGATCTATATTTTCAAAATGGAAGATATGTAATTTTAGACAGA 3464
Db 608 TTATAAATAATTTTTTATAATTTTTTATATATTTATTTTTTTTAAATATATTTATTTTA 667

Qy 3465 CACCTTATAGGACTAGACTATCTTATACAGGCGTCAAAACAGTTGGTGGCTTTATG 3524
Db 668 TTTTATATTTTTTTTTTTTATTAATTTTTTATATATATATATATTTTTTTTTTTTTTTT 727

Qy 3525 AACTATCCTAGTTAAATPACCACATTAATAGTTTCAATTCGGTTAAATCTTTGCACTT 3584
Db 728 NTTTTAAATTTTTTAAATAAATATTTATTTATTTATATATATTTATTTTTTTTATTTAT 787

Qy 3585 ATAAAAATAAATAGCAACAATTTTTTTCTGTGCTGCTGCTTTTATACCAATCTATTTA 3644
Db 788 ATTTTTTTTTTATTTATTTATATTTATTTTATTTTAAATTTTTTTTTTTTTTTTTTTTAT 847

Qy 3645 AGTGGATCGAATTTGGTATTTATCGGTAGCAATATTAATATATATGCTGTTATGGAGA 3704
Db 848 TTTTATTTATTTATTAATATTTTATTTTATTTATTTTTTTTTTTTTTTTTTTTTTTTTTT 907

Qy 3705 TATATAGGTGAAAAATTTGCTTGGATAAAAAAGCTAAATAGTAATTTTGTAAATCTACTT 3764
Db 908 TTTTATTTATTTATTTATTTTATTTTTTTTTTTTTTTTATTTTATTTTAAATTAATTAAT 967

Qy 3765 ATATTTTAAATACGAATGCTTTTACCATGAATTT-----TTGCTGTTTATA 3813
Db 968 AATTTTTTTATTAATTTATTTTATTTTAAATTTTAAATATTTATTTTAAATTTTTTTATTTA 1027
```

```
Qy 3814 ATTCTAGAGAAATCAAGTAAGCAAGCTAGATTTATTTATTTATCAAGGAGTATTTGATTAAG 3873
Db 1028 TTTTATAAAATATATATAAATTTATTTTATTTTATTTTATTTTATTTATTTATTTATTTAT 1087

Qy 3874 TATTAGAAAAACAATATTTTATTTGATATGGAATATCGGAATATTTCAAGTTACGGGAACCTT 3933
Db 1088 TATATTTTTTTTAAATATTTTTTTTTTATTTATTTATTTTATTTATTTATTAATAATATTTT 1147

Qy 3934 GGCTCGGAAGTCATTCAGGCTATATATCATTTTTTTTATAAATCAGGAATAGTTGGTTGA 3993
Db 1148 AAAATATTTATTTTATTTTATANAATAATTTTTTTTTTTTTTTTTTATTTATTTATTTATA 1207

Qy 3994 TTTTACTGATGTTTCTTTTTTTTTTATGTTATAAAAAAGTTATGAGTATTAATCGGGGAAA 4053
Db 1208 TTTAAATTTTTTATATTTTTTTTTTTTTTATTTATAAATAATAATTTATTTATTTATTTT 1267

Qy 4054 CAGCACTATTTTATTTTACATCATTTAGCCATATTTTTCATATATGAACAATAGATCCGA 4113
Db 1268 TATATATAAAT -TATTAATAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1326

Qy 4114 TTATTTATTTATTTAGTACTAT 4134
Db 1327 TATATATTTATTTATTTATTT 1347
```

## RESULT 5

CF238805/c

LOCUS

DEFINITION AGENCOURT 15099447 NICHD\_XGC\_Emb6 Xenopus tropicalis cDNA clone IMAGE:6995950 5', mRNA sequence.

ACCESSION CF238805

VERSION CF238805.1

KEYWORDS EST.

SOURCE Xenopus tropicalis (western clawed frog)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;

Xenopodinae; Xenopus; Silurana.

REFERENCE 1 (bases 1 to 1626)

NIH-MGC http://mgc.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: cgaops-remail.nih.gov

Tissue Procurement: Robert M. Grainger

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLNL4680 row: g column: 21

High quality sequence start: 71

High quality sequence stop: 316.

Location/Qualifiers

1. 1626

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/db\_xref="taxon:8364"

/clone="IMAGE:6995950"

/tissue\_type="neurula"

/dev\_stage="embryo, stages 14-19"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NICHD XGC Emb6"

/note="vector: pCMV-SPORT6.1; Site 1: NotI; Site 2: EcoRV;

Cloned unidirectionally. Primer: Oligo dT. Average insert

size 2.1 Kb. Constructed by Invitrogen. Note: this is a

Xenopus Gene Collection (XGC) library."

ORIGIN







/tissue_type="mixture of kidney and spleen" /clone_lib="MSMg01 Mouse Male BAC Library"									
ORIGIN									
Qy	2604	TGATCGAGCTTTGGAATACACTATTACAAAGATTGGTTGTTATTGATGAGATAAAAAACA	2663						
Db	155	TTTATANAATTTANTTAATTTTAAAAAATTTTAAATATATATATATTTTAAAAAATTTAT	214						
Qy	2664	TGGTCTAAGATAAGATTTCGGTTCTAATTCGGTTTCGCTCCACATGATTTTGGGCAAT	2723						
Db	215	TAATTTAATATAAATTAATTTTAAATATATATTTTAAATATATTTTAAATATTTTAAATAT	274						
Qy	2724	TCT--TTTATCAAAATGAAACGAAACAGCTTATTTTAAAGTAATCTAAATGCCAGAT	2781						
Db	275	TTTAAATATATATTTTAAATTTTAAATAATTTTATTTTATATATATATATTTTAAATTT	334						
Qy	2782	GAACTATTATACAGACAAATTTATAGAAAAATATGAAATTTTCAAAATAGATTACTCAAAT--	2839						
Db	335	NTTAAATTTTATTAATAATTTTAAATAAATTTTATATTTTATTTTATTTATATATTTTA	394						
Qy	2840	ATGGAATTTAAGATATATAAAGTGGAAAAATCAACATCTTCTCCTATTGCTTTACAG	2899						
Db	395	ATATAAATTTTATTTTAAATATATTTTATATAAATTTTATTTTATTTTATTAATAAAT	454						
Qy	2900	ATGATTTCTATTGATGAATTTGCTAAATGCAAGAAATTTAGGTTTTTATTTTCTAGAAAGT	2959						
Db	455	ATTTTATTTTATATATAAATAAATAATTTTATTTTATTTTATTTTATATATATAA	514						
Qy	2960	TAAAAATAGAAAAATCAATCTAAATTTTAAAGAAATTTACTAAAAATAAAATAGTTGAT	3019						
Db	515	TTTATATTTTATTTTATATTTTATATTTTATATTTTATATTTTATATATAATATTTTA	574						
Qy	3020	TTTGTGAGAGTAAGTATGTTTAAATTTTAAATATAGCCCGAAATTTTATTTTAA	3079						
Db	575	TTTTTAAATAATTAATTTTAAATAATTTTAAATAATTTTATTTTATTTTATTTATTT	634						
Qy	3080	GTACTTCTGGTTGATTTTATTTTCCAGACCAAAAGTAGTATTTTATTTTATTTAT	3139						
Db	635	TTATTTAAAAATATTTATTTATTTTATTTTATTTTATTTTATTTATTTATTTTATTT	694						
Qy	3140	GAATTTAATTTTATTCATATAAAAATTTTGAAGCTAAGCTAATATTTAAAAAATGAAT	3199						
Db	695	TTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT	754						
Qy	3200	TTTATTTGTTTATTTATGCTATATTTATGTTTTCAGTAGTCACAGTATGTTGT	3259						
Db	755	TTTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTAT	814						
Qy	3260	TGAATAAAATTTGAAAGATTAT--TTGCAGATTTTACTGCTCCATAAATTTGGATTAT	3317						
Db	815	TTAATTAATTTTATATAATTTATTTATTTTATTTTAAATTTAATAATTTAAAAATATTTATA	874						
Qy	3318	GCAATAATGATTAATAATTTGATTCATTTTAAATAATTTGATTTAAAAAATTTAAAAAT	3377						
Db	875	TATATATTTTATTTTATATAATTTTATTTTATTTTATTTTATTTTATTTTATTTAAT	934						
Qy	3378	AGTATCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTCAAAAT	3437						
Db	935	NTTTATTTTATATATTTTATTTTATTAATAATTTATATATATATTTATTTATTTTATTT	994						
Qy	3438	GGGAAAGATATTTGATTTTATTTTATAGACAGACCTTATAGGACTATCTTTATAACAGGC	3497						
Db	995	TTTAAATATATTTTATTTTATATTTTATTTTATTTTATTTTATTTTATTTTATTTAATNT	1054						
Qy	3498	GTCAAAAACAGGTTGGCTTTATGAACATATCCTACGTTTAAATACCACTACAAATTTATA	3557						
Db	1055	TTATTTTATTAATTTATTTTATTTTATATTTTATATATAATTTTCTTATTTTANTTATTT	1114						
Qy	3558	GTTTCAATTCGTTAATCTTTGCACTTTATAAAAAATAAAATGCAACAATTTTTTTCTTG	3617						

Qy 2028 TTGCGCAGGAGGCCCTACTTTTATGAATTCATTCACAAAGGAAAAAACAATTAT 2087  
Db 87 AA 146  
Qy 2088 TGTTTCCCTAGACAAAAAAGTATGCTGAACATGATAATGATCATCAAGTACAGTTTGTAA 2147  
Db 147 AA 206  
Qy 2148 GAAGAAATTTTCAAGATAATAATATTTTATTTATAGAAAAATATAGATGATTTGTTGAAA 2207  
Db 207 AA 266  
Qy 2208 AATATTTGAAGTTTCTAGCAGCAAACTAATCTTACATCAAAATAAATTTTTTTTGTGAAA 2267  
Db 267 NNN 326  
Qy 2268 GATTAAGCAAAATAGTTGCAAAAAATTTAATGAGGATCAAGAAAAATGAATAAATAAAGAT 2327  
Db 327 NNN 383  
Qy 2328 GCATATTTGATAATGGCTTATCATAAATTTTCTCAGATTTTCTCGAGAGGATACAGAT 2387  
Db 384 AA 443  
Qy 2388 ATTATCATCTCTCTCAGGAGAAATGACACCATAGTTCTCTCAGAAATCCTGTATATTT 2447  
Db 444 AA 503  
Qy 2448 ATTTTAAATATCTCAGGATTTATATGTTGAATTTTACAAAGATGACGACAAATATTAAG 2507  
Db 504 AA 563  
Qy 2508 AAAATAGGATATATGAACGAGTTAAATGTTTACAGATTTATTTCTTAAATATATACAGAAAAA 2567  
Db 564 AA 623  
Qy 2568 CTATTTGATAATGATCTGTTTGAATTTTATTAAGATGATGCGAGCTTTTGAATCTATT 2627  
Db 624 AA 675  
Qy 2628 TACAAGATGTTGTTTATTCATAGATAAATAACATGCTTCAAGATAAGATTTGGTTC 2687  
Db 676 NNN 735  
Qy 2688 TAAATGGGTTTCGCTCCACATGATTTTGTGGCAATCTTTTATCAAAATGAACGAAAC 2747  
Db 736 NNN 795  
Qy 2748 AGCTTATTTTAAAGTAATCTAAATGTCAGATGAACT--ATTATACAGACAAATTATA 2805  
Db 796 ANNNNNNTTNAATAAANNNAANNAATAAATAAATAAATAAATAAATAAATAAATA 855  
Qy 2806 GAAAAATATGAATTTTCAAATAGATATCTAAATATGGAATTTTAAGATATATAAAGTGG 2865  
Db 856 AATATAAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 915  
Qy 2866 AAAAAATCAATCTCTCTCTTATGCTTTTACAGATGATCTTATTTGATGATGCTTAAT 2925  
Db 916 AATAATATNNAANNTTATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 975  
Qy 2926 GCAAGAAATTTAGGTTTTTATTTTCTGAGAAGTTTAAAAATAGAAAAATATCAAAATTT 2985  
Db 976 AATAAANTTATATATATATTTTAAATTTATNTATNTATNTATNTATNTATNTATN 1035  
Qy 2986 AAGAAAAATTTACTAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 3045  
Db 1036 AAAAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1082  
Qy 3046 TATTTAAATATGACCCGGAATTTTATTTTAAAGTACTTCTGTTGATTTATTTTATTC 3105  
Db 1083 TATTAATANATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1142  
Qy 3106 CAGAGCAAAAGTATGTTTATTTTAAATTTTATGAAATTTTATTTTATTCATATAAAT 3165

Db 1143 ATANATAATAAANAATAATAAANAANTAAATTTAAANAATTTAATAAATNTTATTAATAAT 1202  
Qy 3166 TTTTGAATACTAAGCTAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 3225  
Db 1203 TATTNAATAATATNTAANTATTAATAANTATTTATTAATAAATAAATAAATAAATAAATA 1262  
Qy 3226 TATGTTTTTCTCAGTAGTCACAAAGTATGTTTGTGAAATAAATAAATAAATAAATAAATA 3285  
Db 1263 TTTTAAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1322  
Qy 3286 CAGATTTTACTGCTCCCATATTTGGATTTATGCAATAATGCTATTAATAAATTTGATTTTAT 3345  
Db 1323 TAAATAATATNTNTATNTATTTAATAAANTNTAATAAATAAATAAATAAATAAATAAATA 1382  
Qy 3346 TTATAAATAATTTGATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 3405  
Db 1383 TNNNAANTATNTTATTTATNTAATTTTNTATATTTATTAATAAATAAATAAATAAATAAATA 1442  
Qy 3406 TAGGTATATCTGCAATGCTATTAATCAAAATGGGAAGATATTTGATTTTACAGACAGAC 3465  
Db 1443 AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1502  
Qy 3466 ACCTTATAGACTAGACTATCTTATAACAGCGCTCAA-----AACAGGTTGGTGGCTT 3520  
Db 1503 AATAAATAAANAATATTTATTTATNTAATTTATNTAANTNTAANTNTAANTNTAANTNT 1562  
Qy 3521 TATGAATCTATCTTACGTTAAATACCACTACAAATATATAGTTTCAATTCGGTTAACTTTGC 3580  
Db 1563 TTATTTTATATATTTTAAANNANTTAATATANTATNTAATATTTTATTAATAAATAAATA 1622  
Qy 3581 ACTTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 3640  
Db 1623 TTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1682  
Qy 3641 TTTAAGTGGATCGAGATTTGAGTTTATCGCTAGCAATATTAATATATATCGTTGTTATG 3700  
Db 1683 TATANTTTAAATNTAATNTAATNTAATNTAATNTAATNTAATNTAATNTAATNTAATNT 1742  
Qy 3701 GAGATATATA 3710  
Db 1743 AAAAAATAA 1752

RESULT 9  
CL509408/c  
LOCUS SAIL\_H11\_H11.v3 SAIL Collection Arabidopsis thaliana genomic clone  
DEFINITION SAIL\_H11\_H11.v3, genomic survey sequence.  
ACCESSION CL509408  
VERSION CL509408.1 GI:46006728  
KEYWORDS GSS.  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi  
1 (bases 1 to 1758)  
AUTHORS Sessions, A., Burke, E., Presting, G., Aux, G., McElver, J., Patton, D.,  
Dietrich, B., Ho, P., Bacwaden, J., Ko, C., Clarke, J. D., Cotton, D.,  
Bull, D., Snell, J., Miguel, T., Hutchison, D., Kimmerly, B.,  
Mitzel, T., Katagiri, F., Glazebrook, J., Law, M. and Goff, S. A.  
A high-throughput Arabidopsis reverse genetics system  
Plant Cell 14 (12), 2985-2994 (2002)  
22356987  
12468722  
COMMENT Contact: Sessions A  
Applied Trait Genetics  
Syngenta Biotechnology Inc.  
3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA  
Email: allen.sessions@syngenta.com  
ABRC Stock Number CS836276; T-DNA left border flanking sequences of  
Syngenta Arabidopsis Insertion Library (SAIL) lines are available

through the Arabidopsis Biological Resource Center (ABRC).  
Sequences represent a pool of amplified genomic regions and not  
single contiguous sequences.

## FEATURES

source  
1. .1758  
Location/Qualifiers  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/ecotype="Columbia"  
/db\_xref="taxon:3702"  
/clone="SAIL\_811\_H11.v3"  
/clone\_lib="SAIL Collection"  
/note="T-DNA left border sequences were isolated using a  
modified TAIL-PCR strategy"

## ORIGIN

Query Match 2.1%; Score 149.6; DB 9; Length 1758;  
Best Local Similarity 36.2%; Pred. No. 5.7e-13;  
Matches 631; Conservative 0; Mismatches 1086; Indels 27; Gaps 5;  
QY 2134 AGTAGAGTTGTGAAGAAGAAATTTTACAAGATAATAATTTTATTTATATAGAAAATATAGA 2193  
DB 1742 ANTATTAATTTTAT 1683  
QY 2194 TGATTTCTTTGAAAAAATTTTGAAGTTTCTTAACCAAACTAACTTTACATCAAAATATAA 2253  
DB 1582 TANTTAATATTAATATATANTANTTATATATATATATATATATATATATATATATATATAT 1623  
QY 2254 TTTTCTTTGTGAAGATTAACAAATAGTTTGAAGAAATTTTAAATGAGGATCAAGAAAATGA 2313  
DB 1622 NNTANTATTAATAAATATATANTANTANTATATATATATATATATATATATATATATATAT 1563  
QY 2314 ATATAAATAAAGATGC-----ATATTGATTAATGGCTTATCATATATTTTC 2359  
DB 1562 ANNNNAANAATANTANTANNANTANATAATNTNANATAAATAATNTTTATTTATTT 1503  
QY 2360 TCAGATTTTCTGAGAGGGATACAGATATATCATCTCTCTCAGGAGATGCACACCA 2419  
DB 1502 TTTTATTTTATATATATATATATATTTTATATATATATATATATATATATATATATATTT 1443  
QY 2420 TTAGTTCCTTCAGAAATACCTGTATAATTTTAAATATTTCTCAGGATTTATATGTTGAA 2479  
DB 1442 TTTAATTAATATATATTTAT 1383  
QY 2480 TTTACAAAGATGAGCAAAATATAAGAAA-ATAGATATATGAACGAGTTAAAGTTA 2538  
DB 1382 AATAATATATATATATATTTTANNATTTANNAATTTATANAATATANAATANNATATATTT 1323  
QY 2539 CAGATTTATTCCTAATATATCAGAAAACCTATTGATAATGTACTGTTAGAAATTTTATTT 2598  
DB 1322 ATATATTTTNNNTAANTTTTANTTAANNATTTT-ATATTTAATTAATANTTAATA 1267  
QY 2599 AAGAATGTATCGAGCTTTGAATACTATTTTCAAGAGATTTGTTTATTTGATGAATAAAA 2658  
DB 1267 NANAATATNTTANTTATATATTTATATATATANTANTTATATANTANATATATATTTN 1207  
QY 2659 AAACATGGCTCAGAAATGAAGTTGGTTCTAAATTTGGGTTTCGGTTCACATGATTTTGTG 2718  
DB 1206 AATAAATTTTATANNATTTAATAATNTTAAATTTNATNTATTTATTTTATATAT 1147  
QY 2719 GCAATTCCTTTATCAAGTGAACGACAGCTTATTTATTTAAGTAATCTAAATGTCCA 2778  
DB 1146 NTATTAANAANTATATATANAANAANTTATATANTTAANAANTTATTAATNTATTT 1087  
QY 2779 GATGAATTTTATACAGACAAATATAGAAAATATGAAATTTTCAAAATAGATTTCTAAA 2838  
DB 1086 AATAAATTTTATTTATATATATTTATATATATTTATATATANTTTTNTANTNTAAA 1027  
QY 2839 TATGGAATTTAAGATATATAAAGTGGGAAAATAATCAACATCTTCTCTATTTGCTTTACA 2898  
DB 1026 TTTATATTTAATAATNATATATATTTTAAATATATATATTAATTTTATATTAATAAAA 967  
QY 2899 GATGATTTCTATGATGAATTCCTAAATGCAAGAAATTTAGGTTTTTTTATTTTGCTAG--- 2954

RESULT 10  
AG320553  
LOCUS

AG320553 1780 bp DNA linear GSS 02-JUN-2004  
Mus musculus molossinus DNA, clone:MSMg01-106H14.TJ, genomic survey  
sequence.

DB 966 TATTTAATATANAATTTTAAATTTATTTATTTATATAATTTTATATAATTTTNNNAATAT 907  
QY 2955 -AAGTTAAAAATAGAAAATAAATCTAAATCTAAAGAAATTTAATCTAAAGAAATAAATA 3013  
DB 906 AAAATATANATATAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 847  
QY 3014 GTTGATTTTGTGAGAGTAATGATGTTTAAATTTTAAATATTTAAATATGACCCGAATTTTAT 3073  
DB 846 TATTTATTTATTTATTTTAAATTTTATTTTNNNTTTTATNNNNNNNTTTNNNTTNN 787  
QY 3074 TTTTAAAGTACTTCTGCTGATTTATTTTATTTCCAGAGCAAAAGTATGTTATTTTATTAAT 3133  
DB 786 NTTT 727  
QY 3134 TTTTATGAATTTTAAATTTTATTTTCAATAAATAATTTTGAAGAACTAAATATTAATAAAA 3193  
DB 726 NTTT 667  
QY 3194 TGAATTTTATTTGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 3253  
DB 666 TTTNTNTNTTTTTNNNNTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 607  
QY 3254 GTTTGTTGAAAATAAATTTTGAAGAATTTATTTGCAGATTTTACTGCTCCCAATAATTTGGAT 3313  
DB 606 NTTTTNTNTTT 547  
QY 3314 TATTGCAATTAATGTTATTTATTAATTTGTTTCAATTTATAAATATTTGATTTATAAAAAAT 3373  
DB 546 NNTTTNNTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 487  
QY 3374 AATAGTATCTTTTATTTAGTTTTTATTTAGTTTATTTAGGTATATCTGCAATTTGTAATTTCA 3433  
DB 486 TTTTAANNNTTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 427  
QY 3434 AAATGGAAGATATTTGTTATTTTACAGACACACCTTATAGGACTAGACTATCTTATAAC 3493  
DB 426 NNN 367  
QY 3494 AGCGCTCAAAACAAGGTGGTTGGCTTTATGAACCTACCTACGTTAAATACCACTACAAT 3553  
DB 366 ATATATATATATTT 307  
QY 3554 TATAGTTTCAATTCGTTAAATCTTGCACCTTATAAAAAATAAATAAGCAACAATTTTTTT 3613  
DB 306 TTTNNNTTTNTN 247  
QY 3614 CTGTGCTCTGCTTTTATATACCGATCTATTTAAAG--TGGATCGAGAATTCGTAGTTTATC 3670  
DB 246 NTTT 187  
QY 3671 GCTAGCAATTTAATTTATATGCTGTTATGGAGATATATAGTGGGAAAATTTGCTTGGAT 3730  
DB 186 TTTNNNNNTT 127  
QY 3731 AAAAAAGCTAATAGTAATTTTGTAACTACTTATTTATTTTAAATCTGAAATGCTTTA 3790  
DB 126 AAAAAATNTNNNNNNNTTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTA 67  
QY 3791 CCATGAAATTTTGGCTGTTTATATAATTTCTAGAGAATCAAGTAACGAAGCTAGATTATAT 3850  
DB 66 TAAATNAATTTTTTTTTTTTTTTTTTTTTTAAATTAANNANTTTNTTTTAAATTTTNTT 7  
QY 3851 TTAT 3854  
DB 6 TTTT 3

ACCESSION AG320553  
VERSION AG320553.1 GI:47893510  
KEYWORDS GSS.  
SOURCE Mus musculus molossinus  
ORGANISM Mus musculus molossinus  
REFERENCE Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.  
AUTHORS BAC end sequences of Library MSMg01  
TITLE Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1780)  
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.  
TITLE Direct Submission  
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail:hattori@sc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,  
Tel:81-45-503-9111, Fax:81-45-503-9170)  
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC  
library availability, please contact Kunia Abe (abe@rtc.riken.jp).  
Tsukuba Institute, Bio Resource Center,  
The Institute of Physical and Chemical Research (RIKEN) 3-1-1  
Koyadai, Tsukuba, 305-0074 Japan  
phone: 81-298-36-9189, fax: 81-298-36-9199  
e-mail: abe@rtc.riken.jp  
PRIMERS  
Sequencing : TJ  
LIBRARY  
Vector : pBACE3.6  
R.Site 1 : EcoRI  
R.Site 2 : EcoRI  
FEATURES  
source Location/Qualifiers  
1..1780  
/organism="Mus musculus molossinus"  
/mol\_type="genomic DNA"  
/sub\_species="molossinus"  
/db\_xref="taxon:57486"  
/clone="MSMg01-106H14.TJ"  
/sex="male"  
/tissue\_type="mixture of kidney and spleen"  
/clone\_lib="MSMg01 Mouse Male BAC Library"  
ORIGIN  
Query Match 2.1%; Score 148.2; DB 9; Length 1780;  
Best Local Similarity 47.2%; Pred. No. 9.3e-13;  
Matches 712; Conservative 0; Mismatches 776; Indels 20; Gaps 8;  
Qy 2886 TATTGCTCTTACAGATGATTCCTATTGATGAATTCGCTAAATGCAAGAAATTTAGGTTTTT 2945  
Db 247 TATTGATTATATATATATAGATTAATAGATTAATAGATTCCTCTCCAAATATATTAATTTATA 306  
Qy 2946 ATTTGCTAGAAAGTTAAATAATAGAAATTAATCTAAATTTAAAGAAATTTACTATAAAA 3005  
Db 307 TCTTGATATAGTAAAAATTAATNATATATATATATATAATAAATAAAGATATAATAATA 366  
Qy 3006 ATAAATAGTTGATTTGCTGAGATGAATGATGTTAAATTTTAAATATATGACCCGAA 3065  
Db 367 TTAATAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 426  
Qy 3066 TATTTTTATTTTAACTACTCTCGTGGTGAATTTTATTTTCCAGAGCAAAAGTATGTTATT 3125  
Db 427 TACTATTTTAAATTAATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTTTAT 486  
Qy 3126 TTATTAATTTTATGAATTTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 3185  
Db 487 ATAAATTAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 546  
Qy 3186 TTAAAAAAGAAATTTTATGTTTTTATATGCTCTATATATGTTTGTGTTCTAGTAGTC 3245  
Db 547 ATAAATAAT--TTTTTTTTTTTTTATCATCTTTTAAATAATAATAATTTATTTTATTT 604  
Qy 3246 ACAAGTAGTTGTGTGAATAAATTTTGAAGATTTTGTGAGATTTTACTGCTCCCAT 3305  
605 ATAAATTTATAAAATTTATATTTTCTTATAAATTTAATATATTAATA 664  
Qy 3306 ATTTGGATTATTCATAAATGTTATTAATTTTCTTATAAATTTGATTATAAATAAATAA 3365  
Db 665 ATTT--ATATATTTTAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 721  
Qy 3366 AAATTAATAAATAAGTAGTCTTTTTTTAGTTTTTATTTAGTTATATAGTTATATCGCAT 3425  
Db 722 CTATTTATTAATAATTTATTTTAAATTTATATAGTAATAAATAAATAAATAAATAAATA 781  
Qy 3426 ATTTTCCAAATGGGAAGATATTTTATTTTATAGACAGACCTTATAGGACTAGACTAT 3485  
Db 782 TATATAATATTTTATCTAAATTTTATATATATATATAAATAAATAAATAAATAAATAA 841  
Qy 3486 CT-TATAACAGGGCTCAAAACAGGTTGGTCTTATGAACCTATCTCGTTTAAATPAC 3544  
Db 842 ATATAAGAAATTAATTAATTTTATTTTAAATAAATAAATAAATAAATAAATAAATAA 901  
Qy 3545 CACTACAAATATAGTTTCAATTCGGTTAAATCTTTGCACTTATAAATAAATAAATAAATAA 3604  
Db 902 TTTTATATTTATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTAT 961  
Qy 3605 ATTTTCTTCTGCTGCTGCTTTTATACCACTATTTTAAGTGGATCGAATGGTAG 3664  
Db 962 ATTATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1021  
Qy 3665 TTTATCGCTAGCAATATTAATTTATATCTGTTGATGGAGATATATAGTGGAAATTTGC 3724  
Db 1022 ATTTTATTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1081  
Qy 3725 TTGGATAAAAAAGCTAATAGTAATTTTGTAAATCTACTTATTTTAAATACTGCAATT 3784  
Db 1082 TAATATATTTACTTTTAAATTAATACTATAATAATTAATTTTAAATTTTAAATTTT 1136  
Qy 3785 GCTTTACCATGAATTTTGGCTGTTTATTAATCTTAGAGATCAAGTAACGAGCTAGATT 3844  
Db 1137 TATTATATTTTAAATAATTTAATAATAATAATAATAATAATAATAATAATAATAATAA 1193  
Qy 3845 TATTATTTTCAAGGAAGTATTTGATAAGTATTTAGAAAACAATAATTTTATTTGGATAGG 3904  
Db 1194 ATAAATTTATTAATTTTATACATATTTTATATATATATTTTAAATTTTATCTAATTA 1253  
Qy 3905 AATATCCGAATTTTCACTACGGAACTTGGCTCGGAAGTCAATTCAGGCTATATATCAAT 3964  
Db 1254 ATTAACATAATTTATTTTATATATTTTATATATCTACTATCTAATTTTATTTATA 1313  
Qy 3965 TTTTATAAATCAGGAATAGTTGGGTTGATTTTACTGATGTTTTCTTTTTTATGTTAT 4024  
Db 1314 AATATATAAA--ATAAATTTATATTAATATCCATATATATTAATAATTAATAAATAA 1371  
Qy 4025 AAAAAAAGTTATGGAGTTAATGGGAAACAGCACTATTTTATTTTACATCATTAGCCAT 4084  
Db 1372 ACATTACATATAAATATTTTTCNAATTAATAAATATTTTATAAACAATAAATAAATTT 1431  
Qy 4085 ATTTTTCATATATGAAACAATAGATCCGATTTATTTATTTATTTAGTACTATTTCTTT 4144  
Db 1432 ATATTTGTTATTAATTTACTAAGTATTAUCATATATATTTTATTAATAATTAATTA 1491  
Qy 4145 AATAGGTATTTGGAAATTAATAAATTTTAAAGGATATGGAGCAAAAAATGAATGATT 4204  
Db 1492 AATATATATTTTAT 1551  
Qy 4205 TAATTCAGTTTATTTGTTACCAATTTTATATGTCGAAGATTTATCTTGATAAATAATTA 4264  
Db 1552 TATATATAAAAAATAATTTTATATATTTTATATATATATATATATATATATATATAT 1609  
Qy 4265 GTATTTTAAACCAACATATATCTAATTTTAGAGGTTTATCTCGTAAATGATCGAAGTACT 4324  
Db 1610 TAAATATATATTTATATATAATATATATATATATATATATATATATATATATATAT 1667  
Qy 4325 ATGATTTCTGAAAAATTTGCTTAAACTATATATGAAGACGATGGAAGAAATTTAAATTA 4384

```
Db 1668 ATTGTTNATATAAATTTTATTATATACTTATTATAATAATAAAATTAATGATTATATTT 1727
QY 4385 AGAAAATT 4392
Db 1728 ATATAAT 1735

RESULT 11
AG350139/c 1489 bp DNA linear GSS 02-JUN-2004
LOCUS Mus musculus molossinus DNA, clone:MSMg01-146K14.TJ, genomic survey
DEFINITION sequence.
ACCESSION AG350139
VERSION AG350139.1 GI:47923449
KEYWORDS GSS.
SOURCE Mus musculus molossinus
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
BAC end Sequences of Library MSMg01
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 1489)
Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC
library availability, please contact Kuniya Abe (abe@c.c.riken.jp).
Tsukuba Institute, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@c.c.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector : pBACE3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES Location/Qualifiers
source 1..1489
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-146K14.TJ"
/sex="male"
/tissue type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN
Query Match 2.1%; Score 146.6; DB 9; Length 1489;
Best Local Similarity 48.3%; Pred. No. 1.7e-12;
Matches 646; Conservative 0; Mismatches 21; Gaps 8;

QY 2052 TTATGATTCATTATCCAGAGGAAAAACAATTTGTTTCTCTAGACAAAAAGTATG 2111
Db 1455 TTATAATAAATTAATAAGTATATATANATTAAATTAATTAATGNTAATATAAATATT 1396
QY 2112 GTGAACATGTAATGATCATCAAGTAGAGTTTGTGAAGAAATTTTACAAGATAAATA 2171
Db 1395 ATTATATAATATANTGTAAATTTATTATTATTATATAGTATATATATATATATAT 1336
QY 2172 TTTTAT-TTATAGAAAAATATAGATGTTTGTGAAAAAAATTTATGAAAGTTCTTAAGCAA 2230
Db 1335 TATTATAATAATAATAATATAATAATAATAATAATAATAATAATAATAATAATAATAGTA 1276
QY 2231 ACTAACTTACATCAATAATAATTTTTTTTTTGTGAAAGATTAAAAACAATAGTTGAAAAA 2290
```

```
Db 1275 NATTTATAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATT 1216
QY 2291 TTTAATGAGGATCAAGAAAAATGAATAATAAAAAAGATGCATATTTTGATTAATGCTTATCA 2350
Db 1215 TTATAATAATTTNAAAATTAATAAATAATAATAATAATAATAATAATAATAATAATAATAATT 1157
QY 2351 TAATTTTTCTCAGATTTTACTGGAGAGGATACAGATATATATCATCTTCTCTCAGAGAA 2410
Db 1156 TTATATATATATAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1097
QY 2411 TGCACACCAATTAGTTCCTTCAGAAATACCTGTATAAATTTTAAATATTTCTCAGAGTTTA 2470
Db 1096 TTATTTAAAAATTTAAAGTTTATATATTTAAATAATATATATATATATATATATATAT 1037
QY 2471 TATGTTGAAATTTACAAAGATGAGCAAAATATATAAGAAATAATAGGATATATATGACGATT 2530
Db 1036 ATTAATATATAAAGATTTTAAATTTTAAATTTTAAAGANTAAATAATAATAATAATAATT 977
QY 2531 AAATGTTACAGATTTATTCCTAATATATACAGAAAAAACTATTGATTAATGACTGTTTGA 2590
Db 976 AATAAATAAGTAAATATATATAAATTTTAAATTAATAATAATAATAATAATAATAATT 917
QY 2591 ATTTTATTAAGAATGTATCGAGCTTTTGAATACTATTTACAAAGATTGTTGTTTATTGAT 2650
Db 916 ATTTTATTAATAATTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 857
QY 2651 AGAATAAAAAACATGCTCTAAGAATAAGATTTGGTTCTAATTTGGGTTTCGCTTCCCATG 2710
Db 856 AAAAAAAAATAAATAATATATAGATAAATAATAATAATAATAATAATAATAATAATAATAA 797
QY 2711 ATTTTGTGCAATCTTTTATCAATGAAACGAAACAGCTTATTTTATTTAAGTAATCTA 2770
Db 796 ATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 737
QY 2771 AATGTCAGATGAACCTATTTTATACACACAATTTATAGAAAAATATGAATAATTTTCAAAATAGAT 2830
Db 736 TTATATAATTAATTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATT 677
QY 2831 TATCTAAATATGGAATTTTAAGATATATAAAGTGGAAAAAATCAACATCTTCTCTATTG 2890
Db 676 AATAATAATATTATTAAATTTAAATATATAATAATAATAATAATAATAATAATAATAATAA 621
QY 2891 TCTTTACAGATGATCTCTATTGATGAATTCGCTAAATCGAAGAAATTTAGGTTTTTTATTG 2950
Db 620 TATTAATAATTTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 566
QY 2951 CTAGAAAGTTAAAAATAGAAAAATAAATCTAAATTTAAAGAAATTTTACTAAAAATAAATA 3010
Db 565 TAATAAAAAATATAAATTTTATATATATAATAATAATAATAATAATAATAATAATAATAA 507
QY 3011 ATAGTTGATTTTGTGAGAGTAAATGTAATTTTAAATTTTAAATATATGACCCGGAATTTT 3070
Db 506 TTTTATAATTTATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 447
QY 3071 TATTTTAAAGTACTCTCTGTTGATTTTATTTTCCAGAGCAAAAGTATGTTATTTTATT 3130
Db 446 TAATAATAATAATTTTAAATTTATATAAATTTTAA-----AAAAAAAATAATAATAA 390
QY 3131 AATTTTATGAATTTTAAATTTTATTTTCATATAAATTTTGTGAAACCTAAAGCTAATAATAA 3190
Db 389 ATTATAAATTAATTTATATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 330
QY 3191 AAATGAAATTTTATGCTTTTTTATTTATGCTCTATATATATGTTTGTGTTCTCAGTAGTCACAAG 3250
Db 329 AATTTTAAATTTTATTAATAATAA-----ATAATAATAATTATTATATAATAATAATAATT 275
QY 3251 TATGTTTGTGAAATAAAAATTTTGAAAGATTAATTTGCAGATTTTACTGCTCCCAATATTG 3310
Db 274 AAAAAATTTTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 216
QY 3311 GATTATTGCAATAATGCTATTATAAATTTGATTTCATTTATAAATAATAATAATAATAATAAATT 3370
```











```

RESULT 15
CL068807      1242 bp   DNA   linear   GSS 31-DEC-2003
LOCUS        CH216-115B3 Sp6.1 CH216 Xenopus tropicalis genomic clone
DEFINITION   CH216-115B3_ genomic survey sequence.
ACCESSION   CL068807
VERSION      CL068807.1 GI:40524720
KEYWORDS     GSS.
ORGANISM     Xenopus tropicalis (western clawed frog)
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
              Xenopodinae; Xenopus; Silurana.
REFERENCE   1 (bases 1 to 1242)
AUTHORS     Kremitzki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,
              Mardis, E. and Wilson, R.
TITLE       A physical map of the xenopus tropicalis genome
JOURNAL     Unpublished (2003)
COMMENT     Contact: Richard K Wilson
              Genome Sequencing Center
              Washington University School of Medicine
              Email: submissions@watson.wustl.edu
              Insert Length: 175000 Std Error: 0.00
              Seq primer: Sp6 ATTAGTGACACTATAG
              Class: BAC ends
              High quality sequence start: 7
              High quality sequence stop: 57.
              Location/Qualifiers
FEATURES             source
                     1..1242
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-115B3"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN
Query Match      2.0%; Score 143.2; DB 9; Length 1242;
Best Local Similarity 46.9%; Pred. No. 5.5e-12;
Matches 545; Conservative 0; Mismatches 613; Indels 4; Gaps 3;

QY 2484 CAAAAGATGAGCAAAATATAAGCAAAATAGGATATATGAACGAGTTTAAATGTTACAGAT 2543
DB 79 CACAAAATAAAGAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAG 138

QY 2544 TATTTCCCTAATATATATATATATATATATATATATATATATATATATATATATATATAT 2603
DB 139 TTTTATTATTTTATATATATATATATATATATATATATATATATATATATATATATATAT 198

QY 2604 TGTATCAGCTTTTGAATATATATATATATATATATATATATATATATATATATATATATAT 2663
DB 199 AATATTTATATATATATATATATATATATATATATATATATATATATATATATATATAT 257

QY 2664 TGGTCTAAGAAATAGATTTGGTCTTAATTTGGTTTCCCATCATGATTTTGGGCAAT 2723
DB 258 TTTTATATATTTTGGTATTTATATATATATATATATATATATATATATATATATATATAT 317

QY 2724 TCTTTTCAATAGAAACGAAACGAGCTTATTTATTTAAGTAATCTAAATGTCAGATGA 2783
DB 318 TATTTAATTTATATATATATATATATATATATATATATATATATATATATATATATATAT 377

QY 2784 ACTATTTATACAGACAAATATAGAAAATAGAAATTTTCAAATAGATTTCTAAATAGG 2843
DB 378 AATATATATATTTTATATATATATATATATATATATATATATATATATATATATATATAT 437

QY 2844 AAATTTAAGATATATAAAGTGAAAGAAAATCAACATCTCTCCATTTGCTTTACAGATGA 2903
DB 1111111111111111111111111111111111111111111111111111111111111111

```

```

DB 438 AAAAAAAATTTCTTTTAAATTTTAAATATATATATATATATATATATATATATATATATAT 497
QY 2904 TTCTATTGATGAATTCCTAAATGCAAGAAATTTAGGTTTTTATTTTGTAGAAAGTTAAA 2963
DB 498 TATTATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 557
QY 2964 AATAGAAAATAAATCTAAATTTAAAGAAAATTTACTAAAAATAAATAAGTTGATTTTG 3023
DB 558 AATATATTTAATAATAATTTTATATATATATATATATATATATATATATATATATATATAT 617
QY 3024 TGAGAGTAATATGATGTTTAAATTTATTTAAATATGACCCGGAATATTTTATTTTAAAGTAC 3083
DB 618 TTTTATTTATATATATATATATATATATATATATATATATATATATATATATATATATAT 677
QY 3084 TTCTGGTTGATTTATTTTATTCAGAGCAAAAGATATGATATTTTATTTAATATTTTATGAAT 3143
DB 678 TTATATATTTTAAATATATATATATATATATATATATATATATATATATATATATATAT 737
QY 3144 TTAATTTTATTCATATAAATTTTGAAGAACTAAGCTAATATATAAATAAGAAATTTTA 3203
DB 738 ATAAATTAATATATATATATATATATATATATATATATATATATATATATATATATATAT 797
QY 3204 TTGTTTTTATATGTTCTATATATATGTTTGTTCAGTAGTCACAGATATGTTTGTGAA 3263
DB 798 ATATATAATTAATTTTATTTATATATATATATATATATATATATATATATATATATATAT 857
QY 3264 ATAAATTTTGAAGATTTATTCAGATTTTACGCTCCCATATATTTGGGATTTATGCAATA 3323
DB 858 TTTATATATTTTAAATTTTATATATATATATATATATATATATATATATATATATATATAT 917
QY 3324 ATGTATATATATTTGATTTCAATTTATATATATATATATATATATATATATATATATATAT 3383
DB 918 TTTATTTTATATATATATATATATATATATATATATATATATATATATATATATATATAT 977
QY 3384 TTTTGTAGTTTTTGTAGTTTATAGGTATATCTGCATTTGTATATATTTCAAAATGGGAA 3443
DB 978 TATATTTAATAATATATATATATATATATATATATATATATATATATATATATATATAT 1037
QY 3444 GATATTGATTTTATAGACAGACACCTTATAGGACTAGATCTTATATACAGCGCTCAAA 3503
DB 1038 CATTATTAATTTATTTGTTATAATATTTATTTATTA-ATATTCATATTAACATTTTTTT 1096
QY 3504 ACAGGTTTGGTTGCTTTATGACTATCTAGCTTAAATACCACACTACAATATAGTTTCA 3563
DB 1097 ATTAATTTATATATATATATATATATATATATATATATATATATATATATATATATAT 1154
QY 3564 ATTCCGTTAATCTTTCACCTTATATAAATAAATAAATGCAACAAATTTTTTCTTGTGCTT 3623
DB 1155 ATTATATTTTATTTAATTTTAAATATATCTTTTATATAAATAATTTAATATATATATAT 1214
QY 3624 GCTTTTATACCGATCTATTAA 3645
DB 1215 AAATTTAATTTTATTTGTTAA 1236

RESULT 16
AG311015/c
LOCUS        AG311015
DEFINITION   Mus musculus molossinus DNA, clone:MSMg01-093J24.TJ, genomic survey
              sequence.
ACCESSION   AG311015
VERSION      AG311015.1 GI:47883969
KEYWORDS     GSS.
ORGANISM     Mus musculus molossinus
              Mus musculus molossinus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
AUTHORS     Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
TITLE       BAC end Sequences of Library MSMg01
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 2015)
AUTHORS     Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.

```





```

RESULT 18
CL110653
LOCUS
DEFINITION
  CL110653 1594 bp DNA linear GSS 05-JAN-2004
  ISB1-53P23 Sp6.1 ISB1 Xenopus tropicalis genomic clone ISB1-53P23,
  genomic survey sequence.
ACCESSION
  CL110653
VERSION
  CL110653.1 GI:40604288
KEYWORDS
  GSS.
SOURCE
  xenopus tropicalis (western clawed frog)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
  Xenopodinae; Xenopus; Silurana.
REFERENCE
  1 (bases 1 to 1594)
  Kremitzki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,
  Mardis, E. and Wilson, R.
  A physical map of the xenopus tropicalis genome
  Unpublished (2003)
  Contact: Richard K Wilson
  Genome Sequencing Center
  Washington University School of Medicine
  Email: submissions@watson.wustl.edu
  Insert Length: 7500 Std Error: 0.00
  Seq primer: Sp6 ATTTAGGTGACACTATAG
  Class: BAC ends
  High quality sequence start: 390
  High quality sequence stop: 470.
FEATURES
  source
    1..1594
    /organism="Xenopus tropicalis"
    /mol_type="genomic DNA"
    /db_xref="taxon:8364"
    /clone="ISB1-53P23"
    /clone_lib="ISB1"
    /note="Vector: phelobAC11; ISB-1 Xenopus tropicalis BAC
    Library Segment 1"
ORIGIN
  Query Match 2.0%; Score 142.6; DB 9; Length 1594;
  Best Local Similarity 45.2%; Pred. No. 6.7e-12;
  Matches 608; Conservative 0; Mismatches 729; Indels 8; Gaps 2;
  2961 AAAAATAGAAATAAATCTAAATTTAAAGAAATTTACTAAAGAAATAAATAAGTTCATT 3020
  229 AAAAAAGAAAAGAGAGAAACCAAGAGAAAAGCAAAAAAATTTTTTTTTT 288
  3021 TTGTGAGAGTAAGTATGTTTAAATTTTAAATATGACCCGGAATTTTATTTTAAAG 3080
  289 TTGTGTTTTTTTTTTTTTTTTTTTTTTTTTTTATTATTATTATTTTTTTTTTTTT 348
  3081 TACTTCGTGTTGATATTTTATTCAGAGCAAAAGATGATTTTATTTATTTTATG 3140
  349 TTTTATTTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 408
  3141 AATTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 3200
  409 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 468
  3201 TTATTTGTTTTTATTTGTTCTATATTTATTTGTTTTCAGTAGTACAAAGTATGTTGT 3260
  469 TTTATTTTATTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 528
  3261 GAAATAAATTTTGAAGAAATTTTTCAGAGATTTTACTGCTCCCAATTTTGGATTATGCA 3320
  529 TTTTATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 585
  3321 ATATGTTATATATTTCTATTTCAATTTATTAATTTGTTATTAATAAATAAATAAAGT 3380
  586 ----TTTTTTTTTTTTTTTTTATTTATTTTATTTTATTTTATTTTATTTTATTTT 641
  3381 ATCTTTTTTAGTTTTTTAGTTTTTATTTAGTATATCTGCAATGATATTTATTTCAAAATGGG 3440

```

```

Db 642 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 701
Qy 3441 AAAGATATTTGTTATTTTATTTAGACAGACACTTATA-GGACTAGACTATCTTATAACAGGGGT 3499
Db 702 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTCT 761
Qy 3500 CAAAACAAGGTGGTGGCTTTTATGAACATATCCCTACGTTAAATACCACTACAAATATAGT 3559
Db 762 TTATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 821
Qy 3560 TTCAATTCGGTTAATCTTTGCACCTTATAAAAAATAAATGCAACAATTTTTTCTCTGCG 3619
Db 822 TTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 881
Qy 3620 TCTTGCCTTTATATCCGATCTATTTTAAAGTGGATCGAGAAATTTGGTAGTTTATCGCTAGCAAT 3679
Db 882 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 941
Qy 3680 ATTAATTTATATGCTTTGTTATGGAGATATATAGGTGGAAATTTGCTTGGATATAAAAGCT 3739
Db 942 TTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1001
Qy 3740 AATAGTAATATTTGTAATACCTACTTATTTTAAATCTGAATGCTTTTACCATGAAT 3799
Db 1002 TTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1061
Qy 3800 TTTGGCTGTTTATTAATTTCTAGAGAAATCAAGTAACGAAGCTAGATTTTATTTATTTATCAAGG 3859
Db 1062 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1121
Qy 3860 AAGTATTTGATAAAGTATTTAGAAAAAATAATTTTATTTTGGATGAGAAATATCCGAATATTC 3919
Db 1122 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1181
Qy 3920 AGTTACGGAACTTGGCTCGGAAGTCAATTCAGGCTATATATCATTTTATTAATCAAGG 3979
Db 1182 AATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTAT 1241
Qy 3980 AATAGTTTGGGTGATTTTACTGATGTTTCTCTTTTATTTTATTTTATTTTATTTTATTTTATTTG 4039
Db 1242 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1301
Qy 4040 AGTTAAATGGGAAACAGCACATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 4099
Db 1302 TTTTATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1361
Qy 4100 AACAAATAGATCCGATTTATTTATTTATTTATTTATTTATTTATTTATTTTATTTTATTTTATTTG 4159
Db 1362 ATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1421
Qy 4160 TAATATAAATTTTAAAGGATATGGAGACAAAAAATGAATGATTTTATTTTATTTTATTTTATTTG 4219
Db 1422 GTTTTATTTATTTGTAATAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1481
Qy 4220 TACCAATTTATTAATGTCGAAGATTTCTTGATAAATGTTATTAACAGATTTATTTATTAACCAA 4279
Db 1482 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1541
Qy 4280 CATATACTAATTTAGAGGTTATTTCT 4304
Db 1542 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1566

```

```

RESULT 19
CL118721/c
LOCUS
DEFINITION
  CL118721 1608 bp DNA linear GSS 05-JAN-2004
  ISB1-72J8 T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-72J8,
  genomic survey sequence.
ACCESSION
  CL118721
VERSION
  CL118721.1 GI:40612356
KEYWORDS
  GSS.
  Xenopus tropicalis (western clawed frog)
SOURCE

```

ORGANISM	Xenopus tropicalis	Qy	3312	ATTATTGCAATAATGATTAATTAATTTGTTATTCATTTTATAAATATTCATTTATAAATAATTA	3371
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae; Xenopodinae; Xenopus; Silurana.	Db	995	TTTTTATATTAATTTTATATATTTTATATATTTTATATATATTTTATATATATTTT	936
AUTHORS	1 (bases 1 to 1608) Kremitzki, C., Carter, J., McPherson, J., Warren, W., Graves, T., Mardis, E. and Wilson, R.	Qy	3372	AAAAATAGTATCTTTTATAGTTTATAGTTATAGTTATAGTTATAGTTATAGTTAT	3431
TITLE	A physical map of the xenopus tropicalis genome	Db	935	TTAAATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT	876
JOURNAL	Unpublished (2003)	Qy	3432	CAAAATGGGAAGATATGTTATTTTAGACAGACCTTATAGGACTAGACTATCTTATA	3491
COMMENT	Contact: Richard K Wilson Genome Sequencing Center Washington University School of Medicine Email: submissions@wustl.edu Insert Length: 75000 Std Error: 0.00 Seq primer: T7 TAATACGACTCCTATATAGG Class: BAC ends High quality sequence start: 252 High quality sequence stop: 345.	Db	875	TTTTTATATTTTATTTTATATATTTTATTTTATTTTATTTTATTTTATTTTATTT	818
FEATURES	Location/Qualifiers	Qy	3492	ACAGGCGTCAAAACAGGTTGGTGGCTTATGAACATATCCTACGTTAAATACCACTACA	3551
source	1. 1608 /organism="Xenopus tropicalis" /mol_type="genomic DNA" /db_xref="taxon:8364" /clone="ISB1-72J8" /clone_lib="ISB1" /note="vector: pBelOBAC11; ISB-1 Xenopus tropicalis BAC Library Segment 1"	Db	817	ATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT	758
ORIGIN	Query Match 2.0%; Score 140.8; DB 9; Length 1608; Best Local Similarity 44.6%; Pred No. 1.3e-11; Matches 644; Conservative 0; Mismatches 792; Indels 8; Gaps 2;	Qy	3552	ATTATAGTTTCAATTCGGTTAATCTTTGCATCTTATAAAAAATAAAATGCAACAATTTTT	3611
		Db	757	TATTAATTTTCTTATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT	698
		Qy	3612	TTCTGTGCTGCTTTTATACCGATCTATTTAAAGTGGATCGAGAAATTTGCTAGTTATCG	3671
		Db	697	ATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT	638
		Qy	3672	CTAGCAATATTAATATATATCTGTTTATCGAGATATATAGGTGGAATAATTTGCTGGATA	3731
		Db	637	TTAATCTTATTAATAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT	578
		Qy	3732	AAAAAGCTAATAGTAATATTTGTAATCTACTTATTTTAAATACTGAAATGCTTTAC	3791
		Db	577	AAAAAATAATTAATTCGGGTCTTTTATTTTATTTTATTTTATTTTATTTTATTTT	518
		Qy	3792	CATGAAATTTGGCTGTTTATTAATCTAGAGAAATCAAGTAACGAAGCTAGATTATTT	3851
		Db	517	AGTGGTTTTTTTGTGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	458
		Qy	3852	TATCAAGGAAGTATTCATGAATGATTTAGAAAACAATATTTTATTTGATATGGAATATCC	3911
		Db	457	TTTTTTTATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT	398
		Qy	3912	GAATATTCAGTTTACGGGAATCTGGCTCGGAAGTCATTTCAGGCTATATATCATTTTAT	3971
		Db	397	TTATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT	338
		Qy	3972	AAATCAGGAATAGTTCGGTTGATTTTACGATGTTTTTCTTTTTTTTATGTTATAAAAAA	4031
		Db	337	TT	278
		Qy	4032	AGTTATGGAGTTAATCGGGAACAGCACATATTTTATTTTATACATCATTAGCCATATTTTC	4091
		Db	277	TT	218
		Qy	4092	ATATATGAAACAATAGATCCGATTATTTATTTATTTATAGTACTATCTTTTCTCAATAGGT	4151
		Db	217	TT	158
		Qy	4152	ATTT 4155	
		Db	157	TTTT 154	
RESULT 20					
CNS00EVL/c					
LOCUS					
DEFINITION	CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999 Drosophila melanogaster genome survey sequence 17 end of BAC: BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL069706				
VERSION	AL069706.1				
KEYWORDS	GI:4949849				
SOURCE	GSS.				
ORGANISM	Drosophila melanogaster (fruit fly) Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;				





```
Qy 2982 ATTTAAAGAAATATTACTAATAAATAAATAGTTGATTTTGTGAGAGTAAATGATGTTT 3041
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1600 ATTATAGAGTATAGTATATATAAATAATATAAATAATTTTATATAAATAATATTATTT 1541
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3042 AAATTTATTAATATGACCCGGGAATATTTTATTTTAAAGTACTTCTGGTTGATTTATTTT 3101
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1540 ATTATTTATATATATATATATATATATATATATATATATATATATATATATATATAT 1481
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3102 ATCCAGAGCAAAAGTATGATTTTATTAATTTTAAATTTTAAAGATTTTAAATTTTATTA 3161
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1480 ATTATAAATAATATATATATATATATATATATATATATATATATATATATATATAT 1421
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3162 AATTTTGTGAACTAGCTAATATTAATAAAGTAAATTTTATGTTTATTTATGTCCT 3221
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1420 ATGTTTAAATAATTTTATTAATTTGAAATAAATAATTTATGTAATGTAATTTTATTA 1361
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3222 ATATTATGTTTGTTCAGTAGTCACCAAGTATGTTTGTGAAATA---AATTTGAAAGA 3278
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1360 ATATTATATATATATGATTAATTAATTAATTTTATATATATATATATATATATTAAT 1301
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3279 TTATTGCGAATTTTCTGCTCCCATATTTGGGATTAATGCAATATGATTAATTAATTTG 3338
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1300 AGATAGGTAGATATAAATAAATAATATAAATAAGATAATTTATGTTAATTTAT---AATTA 1245
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3339 TATTCATTTAATATATTGATTAATAAATAATTAATAAATAGTATCTTTTATGTTTTTA 3398
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1244 GATAAAATTTAAGTATATATAAATAATATGTAATAAATTTATTAATTTTATTTAT 1185
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3399 GTTTTATAGTATATCTGCATCTATATATTATTTCAAAATGGGAAGATATTGTTATTTTA 3458
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1184 AGTTAATAAATAATATTAATTTATTTATTAATTAATTTAATAATAAATAATTAATTAAT 1125
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3459 GACAGACACCTTATAGACTAGACTATCTTTATACAGCGCTCAAAACAGGTGGTGGC 3518
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1124 AATTAATATTATATAGATATTAATTTAAATAAGTATAATATTAATAATAATTTGA 1065
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3519 TTATGAACCTATCTAGTTAAATACCACATTAATATAGTTTCAATCCGTTAACTTT 3578
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1064 ATTATATATAATTAATTTACTAAAA---TATATAAATAAATAAATAATTAATAATAA 1009
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3579 GCATTTAAAAATAAATAAATGCAACAATTTTTTTTCTGTGCTGCTTTTATACCGATC 3638
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1008 AGATAAATAAATATAGATATAAATAATTTTATTTATTAATTAATAATTTATATAA 949
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3639 TATTAAAGTGCATCGAATGGTATTCGCTAGCAATATTAATTAATATATGCTGTTA 3698
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 948 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATACTG 889
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3699 TGGAGATATATAGTGGAAATTTGCTTGGATATAAATAAGCTAATAGTAATTTGTAATA 3758
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 888 AATATAATTTATATATTTTAAAAATAAATTT---AATAATTTTCTTTATATATATTTGTA 831
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3759 CTACTTATTTTAAATACCTGAATGCTTTACCATGAAATTTTGGCTGTTTAAATCT 3818
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 830 ATATTATATATATATATATATAAATAATTAAGTATATAATTAATAAATAATATAATA 771
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3819 AGAATAACAAGTACGAAGCTAGATTTTATTTATTTATCAAGGAAGTATTTGATAAGTATTA 3878
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 770 AATAAATTTTACTATAATAATAATTTTAAAAATAAATAATTTAATAATAGTTTAAATTT 711
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3879 GAA-----ACATATTTTATTTGGATATGGAATATCCGATATATTCAGTTACGGGA 3929
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 710 ATATAAATATAGTAATATTTTATTTTAAATAAATAAATAAATAAATAAATAAATAAATA 651
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3930 ACTTGGCTCGGAAGTCAATCAGGCTATATATCAATTTTATTAATCAGGAATAGTTGGG 3989
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 650 TATTTAAGGAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 591
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3990 TTGATTTTACTGATGTTTCTTTTTTATGTTTATAAATAAATAAATAAATAAATAAATAA 4049
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 590 ATTATTATATTTAATAAATGATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 531
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Qy 4050 GAAACAGCACTATTTTATTTTATCATCATTTAGCCATATTTTTCATATATGAACAATAGAT 4109
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 530 ATTAAATATATATATTTTAAATTTTAAATTTTATATATAATTTATATAAATAAATAATAT 471
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 4110 CGATTATTATTATATATTAGTACTATTCTTTTCTTCAATAGGTATTTTGGATATAATAAAT 4169
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 470 TATTAAAAATAATTTATAAATAAATAATATAAATTTTAAATAGTATAAATAATAATAAATA 411
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 4170 TTTAAAAAGGATATGGAGACAAATAATGAATGAATTTTCAAGTTATTTGTCACCAATTTA 4229
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 410 TATATAAAAACATATATAATATAATTAATAAATTTTAAATATTTTATTTATATCTTT 351
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 4230 TAATGTCGAAGTATCTTTGATAAATGTATTAACAGTATTTTAAACCAACATATACTAA 4289
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 350 ATATTATATATATATAAATAAATAATTTTACTATATACAGATATAAATAAATAATTT 291
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 4290 TTTTAGAGGTAT 4301
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 290 TGGAGAGGANAT 279
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 22
CK997149/c
LOCUS
DEFINITION
  1313 bp mRNA linear EST 26-MAR-2004
  ip15c02.b1 Brain - Cerebellum library (DOEST8) Canis familiaris
  cDNA clone ip15c02, mRNA sequence.
ACCESSION
  CK997149
VERSION
  CK997149.1 GI:45760829
KEYWORDS
  EST.
SOURCE
  Canis familiaris (dog)
  Canis familiaris
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE
  1 (bases 1 to 1313)
  Balija,V., Nascimento,L.U. and McCombie,W.R.
  ESTs from Canis familiaris cerebellum (dog)
  JOURNAL
    Unpublished (2004)
  COMMENT
    Contact: W. Richard McCombie
    Lita Annenberg Hazen Genome Sequencing Center
    Cold Spring Harbor Laboratory
    PO Box 100, Cold Spring Harbor, NY 11724, USA
    Tel: 516 367 8884
    Fax: 516 367 8874
    Email: mcombie@cshl.org
    Plate: ip15 row: c column: 02
    High quality sequence stop: 1313.
FEATURES
  Location/Qualifiers
    source
      1..1313
        /organism="Canis familiaris"
        /mol_type="mRNA"
        /db_xref="taxon:9615"
        /clone="ip15c02"
        /sex="Unknown"
        /tissue_type="Cerebellum"
        /dev_stage="3 month old normal canine"
        /lab_host="Xl10 Gold"
        /clone_lib="Brain - Cerebellum Library (DOEST8)"
        /notes="Organ: Brain; Vector: pBluescript II SK; Site 1:
        EcoRI; Site 2: XhoI; Library constructed using pBluescript
        XR kit from Stratagene. Cloned cDNA was size selected
        between 1-3 kb. Mark Haskins VMD, PhD, Pathology and
        Medical Genetics, School of Veterinary Medicine,
        University of Pennsylvania, 3800 Spruce Street,
        Philadelphia, PA 19104-6051"
ORIGIN
  Query Match 2.0%; Score 138.6; DB 7; Length 1313;
  Best Local Similarity 48.5%; Pred.No. 2.8e-11;
  Matches 567; Conservative 0; Mismatches 580; Indels 23; Gaps 7;
  Qy 2723 TTCTTTTCAATGAACGAAACAGCTTATTTTATTTAAAGTAAATCTAAATGTCCAGATG 2782
  Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  1310 TTTNNATTTAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1251
```



```
QY 2404 AGGAGATGCACACCATTAGTCTCTTCAGATACCTGTATTAATTTTAAATATCTCA 2463
Db 1193 TAATTTTATATACTATTTATATATTTTAAATATATATATATATATATATTTT 1134
QY 2464 GGATTTATATGTTGAAATTTTCAAAAGATGACGAAAAATATAAGAAAAATAGGATATGA 2523
Db 1133 ATATTTAAATTTATTTTATCTTATATATATATACATATATCTTATATATTTTCAT 1074
QY 2524 ACGAGTTTAAATGTTTACAGATTTTCTCTAATATATACAGAAAACTATTGATAATGACT 2583
Db 1073 TTATATATATATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1014
QY 2584 GTTTAGAAATTTTATAGAAATGATACGAGCTTTTGAATACATTTTACAAAGATGTTGTT 2643
Db 1013 TTATATTTTATATATTTTATTTTATTTTAACTATTTTATTTTATATATATA-----TATTT 959
QY 2644 TATTGATAGATAAAAAACATGGTCAAGATAAGATTTGGTTCTAATTCGGTTTCGCTT 2703
Db 958 ATTAAAAATTAATTTTTTTTTTATATTTTAAATATATATTTTCTTTTATTAATAATATTT 899
QY 2704 CCACATGATTTTGGGCAATTTCTTTATCAATGAAACGAAACAGCCTTATTTTATTAAG 2763
Db 898 TTTTATTTTATTTTAAATATATCTATATTTTAAATTTATTTTATTTTATTTTATTTT 839
QY 2764 TAATCTAAATGTCAGATGAACATTTTATACAGCAATTTATAGAAAAATATGAATTTTCA 2823
Db 838 TTTTATATANATTTTATATATTTTAAATTTTATTTTATATATTTTATTTTCTTTTATATAA 779
QY 2824 AATAGATTTATCTAAATATGAAATTTTAAAGATATATAAGTGGAAAAAATCAACATCTTCT 2883
Db 778 TAAATTTTTTTTTTATAA---ATTTTATATATATATATTTTATATTTTATTTTAAATA 722
QY 2884 CCTATTCTTTACAGATGATCTTATTTGATGAATGCTAAATGCAAGAAATTTAGTGTTTT 2943
Db 721 AATTTTATTTTATTTTATATATATTTTATTTTATTTTATTTTATTTTATTTTATTT 662
QY 2944 TTATTCTGATAGAAATTTAAAAATAGAAAAATTAATCTAAATTTTAAAGAAATTTACTAAA 3003
Db 661 TATTTTAAATATTTTATTAATATATATATTTTATTTTAAATTTTATTTTATTTTATTTT 602
QY 3004 AATATAAATAGTTGATTTTGTGAGAGTAATGATGTTTAAATTTTAAATATTTTAAATATGACCGG 3063
Db 601 AATATATTTTATTAATTTTATTAATTTTATTTTATTTTATATATATTTTATATATATTTT 542
QY 3064 AATATTTTATTTTAAAGTACTCTGTTGATTTATTTTATTTCCAGAGCAAAAGTATGTAT 3123
Db 541 AATTTTTTTTTTTTTTTT---TTTTTAAAAAANAATTTTATTTATAGGGGGGGGGAA 486
QY 3124 TTTTATTAATTTTATGAAATTTAATTTTATTTTCAATATAAAATTTTGGAAAACTAAGCTAA 3183
Db 485 TAAAAATATTTTATTTTATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 426
QY 3184 TATTAAAAAATGAAATTTTATGTTTTTATTTATGTCCTATATATATGTTTGTGTTTCAGTAG 3243
Db 425 AAAAAAATAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 366
QY 3244 TCACAGATGTTTGTGTTGAAATTTTGAAGATTTTTCGAGATTTTTCGCTGCCCA 3303
Db 365 TAATTTTTATTAATTTTATTTTATTTTAAAAAATAAATAATTTTAAATTTTATTAATTTT 306
QY 3304 TAATTGGATTTTGGCAATTAATGATTTATTAATTTTGTATTCATTTTAAATATGATTAATA 3363
Db 305 TTATTTATATATATATATTTTATTTTATTTTAAATTTTAAATATTTTATTTTAAATAAAAAA 246
QY 3364 AAAAAATTAATAATA 3378
Db 245 AAAAAAATTAATAA 231
```

RESULT 24  
AG333983/c  
LOCUS

DEFINITION Mus musculus molossinus DNA, clone:MSMg01-124L03.T7, genomic survey sequence.

ACCESSION AG333983

VERSION AG333983.1

KEYWORDS GI:47907293

SOURCE GSS.

ORGANISM Mus musculus molossinus

Mus musculus molossinus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.

BAC end Sequences of Library MSMg01

Unpublished

2 (bases 1 to 1943)

Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.

Direct Submission

Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Japan

1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,

Tel:81-45-503-9111, Fax:81-45-503-9170)

Clones are derived from the mouse BAC library MSMg01. For BAC

library availability, please contact Kuniya Abe (abe@rtc.riken.jp).

Tsukuba Institute Bio Resource Center, Research (RIKEN) 3-1-1

The Institute of Physical and Chemical Research (RIKEN) 3-1-1

Koyadai, Tsukuba, 305-0074 Japan

phone: 81-298-36-9189, fax: 81-298-36-9199

e-mail: abe@rtc.riken.jp

PRIMERS

Sequencing : T7

LIBRARY : pBACe3.6

Vector : EcoRI

R.Site 1 : EcoRI

R.Site 2 : EcoRI

Location/Qualifiers

1..1943

/organism="Mus musculus molossinus"

/mol\_type="genomic DNA"

/sub\_species="molossinus"

/db\_xref="taxon:57486"

/clone="MSMg01-124L03.T7"

/sex="male"

/tissue type="mixture of kidney and spleen"

/clone\_lib="MSMg01 Mouse Male BAC Library"

## ORIGIN

Query Match 2.0%; Score 137.8; DB 9; Length 1943;

Best Local Similarity 45.7%; Pred. No. 3.6e-11;

Matches 828; Conservative 0; Mismatches 947; Indels 35; Gaps 9;

QY 2095 TAGACAAAAAAGTATGGTGAACATGTAATGATCATCAAGTAGAGTTTGTAAAGAAT 2154

Db 1911 TATTAAGATATAATATGATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1852

QY 2155 -TTTACAGATTAATATTTTATTTATAGAAAATATAGATGATTTGTTGAAAATTA 2213

Db 1851 AATAATTAATATAATAAATAGTAATGATGATGATATATATTAATTAATTAATG 1792

QY 2214 TTGAAGTTCTTAAGCAAACTAACTTTTACATCAATAATAATTTTGTGAAAGATTAA 2273

Db 1791 TATAATAAATTTTGTATTTATATCATATTCGAAAATACATAGATTTATTTATTAAT 1732

QY 2274 AACAAATAGTTGAAAAATTTAATGA--GGATCAAGAAAAATGAATAATAAAAAAGATCAT 2331

Db 1731 AAATGATTATTAATTTATTTTAAATATATGAATAGATATAAAATGCAATAAAATATT 1672

QY 2332 ATTTGATTAATGCTTATCATATTTTCTCAGATTTTCTGGAGAGGATACAGATTTA 2391

Db 1671 ATATTATAGTAACGTTTCATAAGATTTAAATAATTTGTAATTTATATATATAGTATAATGATA 1612

QY 2392 TCATCTTCTCAGGAGGAATGCACACCATTTAGTTCTCTCAGATACCTGTATATATTTT 2451

Db 1611 TAAT-TTATATACTAATTAATAATTTAATGTTTAAATTTAATTAATTAATTAATAGATA 1553

```
Qy 2452 TAAATATTCTCAGGANTTTATATGTTGAATTTTCAAAAGATGAGCAAAATATAAAGAAAA 2511
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1552 TATATATATAGTAGTATATATATATATATATATATATATATATATATATATATATAT 1493
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2512 TAGGATATGACGAGTTAAATGTTGACAGATTATTTCTCTATATATCAGAAAACCTAT 2571
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1492 AGAATTTATATATTTATTTTAAATATCTATAAATGTTCTATTTATATATATGATTAA 1433
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2572 TGATAATGTACTGTTTGAATTTTATTAAGAAATGATCGAGCTTTTGAATCTACTATTTACA 2631
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1432 TATAAATGAATAAATTAATAAATAGTACTTTTCTATAAATTAATAATTAATTAAT 1373
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2632 AAGATTGTTGTTTATGATAGATAAATAAATAAATGCTTAAGAAATGAAATTTGGTCTTAAT 2691
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1372 TATATTATGTCATAATGTAATAAATAAATGTAATGTAATTAATAATAAATAAATA 1314
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2692 TGGGTTTCGCTCCACATGATTTTGTGCAATTTCTTTTATCAATGAAACGAAACAGCT 2751
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1313 ATGATATGAAATTTATATATTTTTGTAAATTAATAATAAATAAATAAATAAATA 1254
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2752 TATTTATTAAGTAAATCTAAATGTCACAGATCA-----ACTATTATACAGACAAT 2801
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1253 AAGATGTTATATAAATATATATATAATAATAATAATAATAATAATAATAATAATA 1194
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2802 TATAGAAAATATGAATTTTCAATPAGATTATCTAAATATGGAATTTTAAAGATATATAA 2861
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1193 AATGAATAAATAAATAATCAAAAAATATTCATAGTATAATTATTAATAAATAATAA 1134
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2862 GTGGAATAAATCAACATCTCTCCATTTGCTTTTACAGATGATTCCTATTGATGATGCT 2921
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1133 TAATAATATATAATTTAATAATATATAAATAAATAAATAAATAAATAAATAAATA 1074
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2922 AAATGCAAGAAATTTAGTGTTTTTATTTGCTAGAAAGTTAAATAATAGAAAATAAATCTAA 2981
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1073 GTAATTTATATTAAATGTAATAAATTAATGTAATAATTAATTAATTAATAATAATA 1014
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2982 ATTTAAAGAAATTTACTTAAATAAATAAATAGTTGATTTTGTGAGAGTAATGATGTTT 3041
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1013 ATATTAAATAATATAAATAAATTAATAATTTATATAATAAATAAATAAATAAATAA 954
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3042 AAATATTATTAAT---ATGACCGGAATATTTTATTTTAAAGTACTTCTGCTGATTAAT 3098
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 953 TATATATTATTAATGAAATAATAATCATCTTTAAATAATAATAATGTAATAATTCATTTAA 894
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3099 TTTATCCAGAGCAAAAGTATGTTTATTTATTTTATGAAATTTAATTTTATTTTATCAT 3158
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 893 TAGATAATAAATAAATAAATAAATTAAGATGACTTTTCA-----GANATGATTAGAA 840
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3159 ATAAAAATTTTGAAAACTTAAGCTAATATTAATAAATAAATAAATTTTATGTTTTTATG 3218
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 839 ATATTATATTAGATAGATATAGTAATTTGTAATAAATAGTTAATTAATATTGATTTTA 780
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3219 TCTATATATGTTTGTGTTTCAAGTACGACAAAGTATGTTTGTGAAATAAATTTTGAAGA 3278
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 779 TATATATAATAAAGTATTTTATATATAATAATAATAATAAATAAATAAATAAATA 720
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3279 TTATTTCAGATTTTACTGCTCCCATATTTGGATTTATGCAATAAATGATTAATAATTG 3338
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 719 ATGTTGAGAAATATATAAATTTTAAATATATATATTGAAATAAATAAATAAATAA 660
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3339 TATTCAATTTAATAATATTGATTATAAATAAATAAATAAATAAATAAATAAATAAATA 3398
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 659 AATTATATTAGATATTGTTAAACATAATATAAATAAATAAATAAATAAATAAATA 600
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3399 GTTTTATTTAGGTATCTGCAATGTTGATATTATTCAAAATGGAAGAATTTGATTTTAA 3458
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 599 ATATAAATACA-----ATAATAATAAATAAATAAATTTTATTAATTTTATGTTTAA 549
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3459 GACAGACACCTTATAGGACTAGACTATCTTATACAGCGCTCAAAACAGGTTGGTGGC 3518
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 548 TATATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 489
```

```
Qy 3519 TTTATGAACATCTACCTACGTTAAATACCACTACAATATTAGTTTCAATTCCTGTTAATCTTT 3578
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 488 TTAATTTATTTTAAATATTTATATATTTAAATATTATGAGTATAATATAAATTTATTATAATA 429
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3579 GCACCTATAAAAAATAAATGCAACAATTTTTTTTCTGCTGCTTTTATACCGATC 3638
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 428 AATATTATATTTATATATAGGAAAAAAGAATTTAAATTAATTTGATTATATATATAT 369
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3639 TATTTAAGTGCATCGAATTTGGTAGTTTATCGCTAGCAATATTAAATTTATATGCTTTGTTA 3698
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 368 TTAATTTATTAATTAAGATTAAATTAATATAATAATAATAATAATAATAATAATAATC 309
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3699 TGGAGATATATAGGTGGAAATTTGCTTGGATAAAAAGCTAAATAGTAATTTTGTGAATA 3758
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 308 AATAAATAAGTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 249
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3759 CTACTTATTTTAAATATCTGA--ATTGCTTTTACCAGTAATTTTGGCTGCTTTTATATTT 3816
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 248 ATTATTATTAATAATAAATATTATTAAATTTAAATAATATAATAATAAATAAATAAATA 189
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3817 CTAGAGATCAAGTAACGAGCTAGATTATTATTATCAAGGAAGTATTGATAAAGTAT 3876
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 188 CTTTAAAGAAACACACACACACACACACACACACACACACACACACACACACACAC 129
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3877 TAGAAAAACA 3886
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 128 AAAAAAANA 119
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 25
LOCUS CL082000
DEFINITION CH216-165P18 RM4.1 CH216 Xenopus tropicalis genomic clone
CH216-165P18, genomic survey sequence.
ACCESSION CL082000
VERSION CL082000.1 GI:40537913
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1981)
AUTHORS Kremitzki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,
Mardis, E. and Wilson, R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 175000 Std Error: 0.00
Seq primer: RM4 ctcaaggggcctcgtcgagc
Class: BAC ends
High quality sequence start: 265
High quality sequence stop: 497.

FEATURES
source 1..1981
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/strain="Nigerian frog"
/db_xref="taxon:8364"
/clone="CH216-165P18"
/sex="male"
/cell_line="Stock 248 F7A2, inbred N7"
/clone_lib="CH216"
/notes="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
BAC library"

ORIGIN
Query Match 2.0%; Score 137.6; DB 9; Length 1981;
Best Local Similarity 43.4%; Pred. No. 3.8e-11;
Matches 674; Conservative 0; Mismatches 876; Indels 4; Gaps 2;
```

[illegible]

Search completed: April 29, 2005, 15:55:26  
Job time : 19346 secs